

20040
74P.

The Chemical Basis for the Origin of the
Genetic Code and the Process of Protein Synthesis

Final Report

To

National Aeronautics and Space Administration
Grant Number 01-010-001

From

The Department of Biochemistry
University of Alabama at Birmingham
University Station
Birmingham, Alabama 35294

May 16, 1990

(NASA-CR-186590) THE CHEMICAL BASIS FOR THE
ORIGIN OF THE GENETIC CODE AND THE PROCESS
OF PROTEIN SYNTHESIS Final Report (Alabama
Univ.) 74 p CSCL 06C

N90-22205

Unclas
H1/51 0279340

In the Interest of efficiency, this report is submitted as two previously submitted documents which together summarize the work accomplished on the grant for the entire period. One is a proposal which was submitted September ,1987 and summarized the data on the grant NGR 01-010-001 up to that date. The other is the annual report for the project covering the work from December,1987 through November,1988. The documents are in reverse chronological order so that the most recent data is seen first.

The Chemical Basis for the Origin of the
Genetic Code and the Process of Protein Synthesis

Annual Report

To

National Aeronautics and Space Administration

Grant Number 01-010-001

From

The Department of Biochemistry
University of Alabama at Birmingham
University Station
Birmingham, Alabama 35294

December 16, 1988

Abstract

We continue to be pleased with progress on a model for the origin of protein synthesis. The essential features of the model are that 5'-AMP and perhaps other monoribonucleotides can serve as catalysts for the selective synthesis of L-based peptides. A unique set of characteristics of 5'-AMP is responsible for the selective catalysis and these characteristics are described in detail.

The model involves the formation of diesters as intermediates and selectivity for use of the L-isomer occurs principally at the step of forming the diester. However, in the formation of acetyl phenylalanine-AMP monoester there is a selectivity for esterification by the D-isomer. Data showing this selectivity is presented. This selectivity for D-isomer disappears after the first step.

NMR studies have confirmed the identity of all four of possible diesters of acetyl-D- and -L phenylalanine with 5'-AMP.

Some new data is also presented on the possible structure of the 2' isomer of acetyl-D-tryptophan-AMP monoester. The data using fluorescence and NMR show the Trp ring can associate with the adenine ring more strongly when the D-isomer is in the 2' position than it can when in the 3' position. These same data also suggest a molecular mechanism for the faster esterification of 5'-AMP by acetyl-D-phenylalanine.

We have also studied the HPLC elution times of all four possible acetyl diphenylalanine esters of 5'-AMP, these peptidyl esters will be products in our studies of peptide formation on the ribose of 5'-AMP.

Other studies were on the rate of synthesis and the identity of the product when producing 3'Ac-Phe-2'tBOC-Phe-AMP diester. HPLC purification and iden-

tification of this product have been accomplished. We now will proceed to the study of kinetics of peptide formation from this diester and perhaps the glycine diester also.

We have finalized data on the hydrolysis of phenyl-3-propionic acid-AMP anhydride as a function of pH.

Our continuing studies on the evolution of tRNA are also discussed.

EXPERIMENTAL

Much of the work in the past year has concentrated on the behavior of 5'-AMP in esterification reactions with amino acids and the properties of the amino acid esters of 5'-AMP. During the course of the year, and after several characteristics of the 5'-AMP had been elucidated, it became obvious that 5'AMP, and probably 5'-GMP, could serve as a catalyst for the selective synthesis of L-based peptides.

The several properties of 5'-AMP and its aminoacyl esters that allow this selective synthesis of L-based peptides were discovered in a sequence in which one finding led to the next idea, etc. The first finding of importance was that, even though all amino acids constantly migrate between the 2' and 3' positions, glycine and all L-amino acids distribute ~67% to the 3' position and ~33% to the 2' position, D-amino acids distribute to the 3' position in a manner inversely related to their hydrophobicities (Lacey et al, 1988). This property, we proposed, was one of the reasons for the origin of the use of L-amino acids in protein synthesis and the exclusive use of the 3' position of the AMP residue at the end of tRNA for formation of the peptide bond.

Comparison of Ac-D- and L-Phe esterification of 5'-AMP

Rather unexpectedly some early results suggested that esterification of 5'-AMP with a racemic mixture of Ac-D- and L-Phe-imidazolide gave a mixture of esters which was 65% D-isomer and 35% L-isomer. This was surprising since other results suggested that separately the two isomers seemed to react with 5'-AMP at about the same rate. However, we soon found that preparations of Ac-Phe-imidazolide would racemize after preparation. What this meant was that although we would start the reaction with one pure isomer or the other we would eventually end up with a mixture of both.

To get a clearer picture of what was happening we decided to prepare separately Ac-D- and L-Phe-imidazolides and store them at room temperature, removing samples from each periodically and reacting them with 5'-AMP for five minutes, stopping that reaction and assaying the esterification products by HPLC for the content of D- and L-esters. As can be seen in Fig. 1, where we have plotted the percent D-isomer as a function of time, the initial D-sample went down to ~65% D and the initial L-sample went up to ~65% D after about 6 hours. The fact that both samples came to the same % D suggests that we have a racemic mix in the reaction but that the esterification with D-isomer is simply proceeding at a rate $65/35 = 1.86$ times as fast as with L-isomer. This observation is somewhat confusing in that in the origin of living systems this preference for D-isomer was somehow overcome. Anyway it represents a sort of obstacle in the evaluation of an all L-isomer protein synthesizing system. However, this obvious advantage of the D-isomer was only operative in the formation of the monoester.

Diester Formation

During the course of these investigations, it became obvious that, not only

could one produce monoesters of the ribose of 5'-AMP, but diesters also. This finding was first reported in our December 11, 1987 report to NASA and included di-acetylglycine and diester of phenyl-3-propionic acid and 5'-AMP. We also reported in that same report that esterifications of 5'-AMP to form the monoester seemed to occur exclusively at the 2' position ie. amino acids enter at the 2' position even though they immediately begin migrating back and forth between the 2' and 3' positions.

The fact that diesters of amino acids could form suggested that, if one of those amino acids had a free amino group, then a peptide bond could form. That would then leave one position on the ribose free for bringing in another free amino acid, forming another peptide bond and continuing the process indefinitely. But the fact that monoester formation seemed to occur exclusively at the 2' position introduced another possibility, but only if diester formation ie addition of the second amino acid, occurred at the 2' position also. If that were true, then there should be some rate advantage for synthesis of L-based peptides by this mechanism. The advantage would come from the fact that all L-amino acids, after monoester formation, distribute preferentially to the 3' position, leaving the 2' position open for formation of the diester. Hydrophobic D-amino acids, on the other hand, tend to remain more in the 2' position and so block, to a greater extent, the entry of the second amino acid.

Comparative Rates of Diester Formation

To determine whether such an advantage could be realized, we carried out experiments with formation of diester from preexisting Ac-D-and-L-Phe-AMP monoester. These monoesters (0.009M) were reacted at 0 °C in a aqueous solution

with a ten fold excess of either Ac-D- or L-Phe imidazolide and samples were removed periodically and assayed by HPLC for the amount of diester formed and the amount of monoester remaining. The progress of the reaction is recorded in Fig. 2. The amount of diester formed correlated directly with the amount of monoester disappearing (Fig. 3). NMR was done on each diester product, proving its identity as diester. Pseudo first order plots for the disappearance of the monoesters are in Fig. 4 and the rate constants are in Table I. These experiments show quite conclusively that whether adding a D- or an -L amino acid as the second amino acid, the rate of adding the second amino acid to a preexisting L-amino acid monoester is about three times faster than to a preexisting D amino acid monoester. These differences we believe are principally due to the fact that the Ac-L-Phe monoester distributes 67% to the 3' position but the Ac-D-Phe monoester distributes only 50% to the 3' position. The advantage to the L-system is more than one would predict simply from the relative amount of open 2' position so other factors must be operating to enhance the reactivity of the preexisting L-monoester.

While one can infer from the above results, that the 2' position must be the one involved in adding the second amino acid, we needed to prove whether this is so.

Diester Forms at the 2' Position Also

To do this, we synthesized a diester in which the two esterified acids were different. First we produced the monoester of butyric acid and purified it by HPLC. We then reacted that monoester of AMP with a five fold excess of Ac-gly imidazolide. The resulting diester was also purified by HPLC and then analyzed by NMR. In Fig. 5 we see a comparison of the NMR spectrum of this diester with

a butyric acid monoester of 5'-AMP which we had obtained earlier. This comparison allows us to determine the location of the butyric acid residue in the diester because we can identify the butyric acid signals from the 2' and 3' ester in the monoester. By integrating the 2' and 3' ribose protons just to the left of the H₂O peak in the lower spectrum we can determine the amount of 2' and 3' butyric acid ester. These results show an excess of the 3' ester. Looking at the signals from the butyric acid chain upfield we can see two sets of signals for each kind of proton. Looking for example, at the terminal CH₃ protons at about 0.9 ppm, there are two sets, one larger than the other. Because we know the 3' ester is in excess, we can identify the larger set as being 3' ester signals and the smaller as 2' ester signals. Comparing the upper spectrum of the diester with the lower spectrum we can then conclude that the butyric residue in the diester is in the 3' position. From this we can infer that the second ester produced was added at the 2' position almost exclusively.

This very nice result allows us to propose, quite seriously, a model for the preferential synthesis of L-based peptides on the ribose of 5'-AMP. The model is presented in Fig. 6 using the synthesis of a tripeptide Ac-PhePhePhe as our example. The step of peptide bond formation has still only been studied in a cursory sort of way.

Synthesis of 3'Ac-Phe-2'tBOC Phe-AMP

We then began a study of the peptide bond forming step. First we produced Ac-Phe AMP monoester and purified it using a C₁₈ Sep-pak unit which allowed the separation from 5'-AMP and from any diester formed. Taking this material, we then reacted it in DMF with a ten fold excess of tBOC-L-Phe imidazolide at room temperature. Samples were taken periodically and analyzed by HPLC using

50%CH₃OH;0.05 phosphate and a phenyl reverse phase column. As time passed the Ac-Phe-AMP monoester peak at 3.8 minutes was disappearing, as shown in Fig. 7. We presume, because the major new peak is the one at 18 minutes, it is the desired product 3'Ac-L-Phe 2' tBOC-L-Phe AMP diester. In order to prove that we collected several micromoles of this material and will now subject it to NMR in comparison Fig. 8 with the monoester of tBOC-L-Phe-AMP in a manner similar to the comparison in Fig. 5.

Now that we have this product produced and as soon as it is identified, we can proceed to the next step of peptide bond formation. This will require first the removal of the tBOC group with trifluoroacetic acid and secondly formation of the peptide bond by incubating at a high enough pH so that the L-amino group of the Phe in the 2' position is unprotonated.

HPLC Study of AcPhePhe-AMP Monoesters

Should peptide bond formation occur, our product will be Ac-Phe-Phe-AMP monoester. Also if we carryout all possible isomeric combinations we will have four different products. Because we want to follow the progress of these reactions with HPLC we needed to find the HPLC conditions and retention times of the various products. Accordingly, we purchased all four Ac-PhePhe isomers (LL,DD,DL and LD), prepared the 5'-AMP monoesters and determined their elution times using a phenyl reverse phase column with 30% CH₃OH;0.05 phosphoric acid at 1.0ml/min. The data are in Table II. Interestingly the homopeptidyl (DD and LL) esters are eluted much faster than the heteropeptidyl esters. We also determined the equilibrium 2'-3' distributions of these compounds. In all cases, the peptides having L amino acids at the carboxy end (and esterified to the 5'-AMP) distributed ~60% to the 3' position and those having D- amino acids

at the carboxy end had about 50% in the 3' position. So, with peptides, as with single amino acids, the 2'-3' equilibrium depends on the isomeric identity of the amino acid directly esterified to the 5'-AMP.

Having the ability to separate these products with HPLC now allows us to proceed with the study of peptide bond formation. We are synthesizing and purifying (by HPLC) large amounts of 3'AcPhe2'tBOCPhe-AMP for the peptide synthesis study. Obviously, we will have to study all four possible combinations of D-and L-isomers. We expect to see some differences in the rate of peptide bond synthesis from the various combinations.

Structure of 2'-AC-D-Trp-AMP Monoester

In our earlier studies of the 2'-3' distributions of various D- and L-amino acids when esterified to 5'-AMP (Lacey, et al 1988), we showed that very hydrophobic D-amino acids, eg. Trp, tended to favor the 2' position. The reason for this seemed to be that, when hydrophobic amino - or carboxylic acids are in the 2' position, they can interact hydrophobically with the hydrophobic adenine ring. NMR studies have supported this idea. We have now carried out a study of the fluorescence of the separate 2' and 3' esters of Ac-D-Trp AMP. The ester was prepared by our normal procedures and the 2' and 3' forms separated by HPLC at pH 2 using 20% CH₃OH;0.05 phosphoric acid at 1.0ml/min. Under these conditions the 2' and 3' esters are well separated and because of the low pH they do not re-equilibrate to form a mixture of the two forms.

Trp is a fluorescent material and can be excited at about 290nm and the fluorescence peak is at about 350nm. After collecting the two separate isomers,

they were adjusted to exactly the same concentration and the fluorescent spectra run with excitation at 290nm. There was no difference in the fluorescence of the two esters. What we had hoped to see was some quenching of the fluorescence of the 2' ester because of hydrophobic interaction of the Trp residue with the adenine residue. Changing the environment of a fluorescing compound can sometimes cause this. However, the absorption spectrum of 5'-AMP and the emission spectrum of Trp do not overlap, so it is perhaps unreasonable to expect an effect. We then went through the same experiment only using a fluorescent analog of 5'-AMP, 5'-ethenoadenosine monophosphate. The absorption spectrum of this compound does overlap the emission spectrum of Trp. In this case there was a great difference in the fluorescence spectrum of the 2' and 3' esters (Fig. 8) when exciting at 275nm. Even though this result is completely supportive of the idea that the Trp residue in the 2' position is associating with the adenine ring, the spectra cannot be interpreted quantitatively. This is because both the ethenoadenosine and the Trp are fluorescing.

NMR done on the 2' and 3' Ac-D-Trp esters of 5'-AMP also supported the interaction in the 2' ester. This result showed a greater upfield shift of adenine ring protons in the 2' ester. Such upfield shifts are caused by association of the adenine protons with the ring of 5'-AMP.

It does appear quite certain then that hydrophobic D-amino - or carboxylic acids when esterified to the 2' position of 5'-AMP do interact with the adenine ring more strongly than when in the 3' position.

This observation also may explain why, as reported earlier in this report, hydrophobic D-amino acids seem to react faster with 5'-AMP than do hydrophobic

L-amino acids. This would be explained by a stabilization of the reaction intermediate by the D-amino acid approaching the 2' position, which is the esterification position, and being held by the Trp-adenine association. The L-amino acid approach would not be so facilitated.

Hydrolysis of Phenyl-3-propionic-AMP Anhydride

In our annual report for 1987 (Dec. 11, 1987), we reported on hydrolysis data for the compound phenyl-3-propionic-AMP anhydride (See Fig. 8 that report) which is a deamino analog of phenylalanine-AMP anhydride. The data there showed the hydrolysis rate constants of this compound plotted versus pH give a U-shaped curve rather than a W shape as had been earlier reported. We have now essentially repeated the entire experiment and included the effect of buffer concentration on the hydrolysis. At each pH, the rate constant for hydrolysis was determined at three buffer concentrations, 0.0125, 0.025 and 0.05 M. In no case was there a significant buffer effect on the hydrolysis. The data are plotted in Fig. 9.

Evolution of tRNAs

Demonstrating evolutionary relationships between RNAs through mutual relationships with a third sequence

In previous work we described matching regions between the sequences of transfer RNA (tRNA) and ribosomal RNA (rRNA)). These matches are too frequent to attribute to chance. In addition, the frequency of matches found in interspecific searches are not significantly different from that found in intraspecific searches. We take this as indicative of true homology (i.e. common ancestry).

Simple pairwise comparisons between macromolecular sequences may not uncover relationships which may be revealed by their mutual relationships with a third molecule. For example, corresponding sites of the Euglena chloroplast and yeast small subunit rRNAs show only an insignificant match with each other but show extensive matches with tRNA^{arg} from *Euglena* chloroplast.

1171 UGGGCUACA CACGUG CUAC 1189 *Euglena chloroplast* 16S rRNA
 ::::: ::::: ::::: ::::: ::::: E = .0560
 16 UGGACUAGAGCACGUG GCUACGAACU ACGGAGUCAG 51 *Euglena* tRNA^{ser}
 ::::: ::::: ::::: ::::: ::::: E = .0097
 1447 GCACGCGCGCUAC A CUGACGGAGCCAG 1473 yeast 18S rRNA

for triad. E = 4.8×10^{-9}
 (1230 threesomes represented by 5 rRNAs and 123 tRNAs were compared).

The configuration including the three molecules has a very low probability of occurrence by coincidence. The low probability translates into a low expected value that takes into consideration the number of opportunities for a given match to be found along and among all of the molecules searched.

The position of the rRNA sequences reflect an overlap of homologous regions based upon the secondary structure of the molecules. Thus the tRNA shows a match beginning with the Euglena chloroplast 16S rRNA and continues by matching a contiguous portion of the yeast 18S rRNA.

These results might be explained as the result of recombination or gene conversion. Presumably the evolutionary history of these RNAs included a large number of such events that resulted in the scattering of matching segments, many of which are large enough to be recognized. The preservation of the segments must reflect constraints on both mutation and recombination, probably because of interlocking functions that would require compensating changes at multiple sites.

Distribution of Matching Regions Among *E. coli* tRNA

Shortly after the first tRNA sequences were determined it was noticed that certain bases tended to be present at specific sites. With the elucidation of more sequences, it became clear that there are certain invariant or semi-invariant bases common to most tRNAs. While some portions of the tRNA molecule are almost universally conserved, other regions are more or less variable. The question arises as to whether any phylogenetic information remains in contemporary tRNA sequences or if the view held by many is correct that tRNA sequences have mutated to equilibrium and no phylogenetic information may be derived from them. Some workers have used specific tRNA sequences from several different organisms to construct phylogenetic trees. The trees based on tRNA sequence comparison mirror phylogenetic trees produced using other parameters. The similarity of trees produced by different methods indicates that tRNA sequences have not lost all their phylogenetic information.

Recently we have described extensive matching regions between *E. coli* tRNAs (especially those between tRNAs for Ile - Val and Ile - Lys encoding tRNAs). The expected values of these matches are so low, and the proportions of the molecule covered by the matches so large (up to 92%) that these matches are considered to represent true homologies.

What remains to be determined is how common and widespread these matches are. To address this question of frequency and quality of matches between tRNAs, an exhaustive comparison was made between all reported tRNA sequences from *E. coli*.

Thirty-five *E. coli* tRNA sequences were compared with each other and matching sequences identified using the Los Alamos routines described by Goad

and Kanehisa (1982). E , the expected number of such matches, per search, was calculated for each match using the formulae of Goad and Kanehisa (1982) as modified by Staves et al. (1987).

The Los Alamos routines were used nevertheless for the following reasons. It is generally accepted that tRNAs all had a common origin (Jukes, 1966). The question which remains today is whether traces of this common ancestry can be found in regions of the tRNAs other than the universally conserved sites. Such traces might make possible the construction of phylogenetic trees for certain tRNAs. Since the genetic code is essentially universal, tRNAs must have had their separate adaptor functions in the translation process since the time of the Last Common Ancestor, selection has worked to make each tRNA molecule highly efficient and specialized for its role in protein synthesis. Part of this specialization involves interaction with specific amino acids and a specific charging enzyme: an aminoacyl tRNA synthetase. This specialization must have required mutation away from the ancestral tRNA sequences. Different tRNAs have different aminoacyl tRNA synthetase recognition sites (Hou and Schimmel, 1988) and thus must have been modified in regions corresponding to different sites on the ancestral tRNA. These modifications and others required to optimize the function of each tRNA may have resulted in significant changes over much of the tRNA sequences. Since we do not know in most cases which portion of the sequences have been modified, local searches (i.e., searches which look for the best matching regions within the two sequences) for matching regions were used. The E value was used as a method of comparison because it takes into consideration several factors including: base composition of the sequences; numbers of gaps, insertions/deletions, and mismatches in the matching regions; length of the

matching region; and length of the two sequences being compared. We feel this gives a more accurate indication of relatedness between the two molecules than simply reporting the frequency of matching bases.

Table 3 shows a matrix containing E for each comparison. 10^{-4} was chosen as our cut-off value. It is clear that most comparisons yield a match of at least this quality. In fact, only 17% of the searches yielded no match $E < 10 \times 10^{-4}$. Figure 10 shows the distributions among non-isoaccepting tRNAs. Very extensive matches are also relatively rare. Matches of $E < 10^{-12}$ occur 54 times, comprising 10.5% of the searches.

The characteristics of the curve shown in Fig. 10 can be explained by the sites on the tRNA sequences where the matches occur. The peak at $E = 10^{-4} - 10^{-6}$ is a result of the fact that in most comparisons between tRNAs a match is found in either the D or T ψ C loop or stem regions (generally bases 10 to 25 and 49 to 66, respectively). A second small peak at 10^{-8} to 10^{-11} is caused mainly by searches that yield matches in both the D and T ψ C arms.

The data presented in Table 3 and Figure 10 indicate that phylogenetic information pertaining to the evolution of the tRNAs themselves may still be found in their sequences. It is anticipated that the information presented here will provide much useful information for future studies of the phylogenetic relationship among tRNAs.

Evolution of tRNA^{ser}

Serine-encoding tRNAs are found in two groups in the genetic anticode; those with anticode XGA, and those with anticode ACU or GCU. The two groups of serine-tRNAs are unusual in that their sequences differ from each other to a much greater extent than commonly occurs between iso-accepting tRNAs. These differences were first exposed in an examination of serine tRNAs from the archaeobacterium, *Halobacterium volcanii* with a collection of tRNAs from widely divergent organisms. In addition to finding matches that occur at the same site on both tRNAs, matches have been found to occur at dissimilar sites on the different molecules. For example:

<i>H. volcanii</i> tRNA ^{ser2} (NGA)	26	GCGGUAGAUUC	36
		•••••••• •••	
<i>E. coli</i> tRNA ^{asp1} (GUU)	4	GCGGUAG UUC	13

These matching regions have been termed oblique matches. In many cases, when an oblique match is found, a corresponding "straight" match is also found.

<i>H. volcanii</i> tRNA ^{ser2} (NGA)	35	UCGAAAUCUAC	45
		•••••••• •••••	
<i>H. volcanii</i> tRNA ^{met1} (CAU)	56	UCGAA UCUAC	65
	56	UCGAAUCUCACC	67
		•••••••• •••	
	56	UCGAAUCU ACC	66

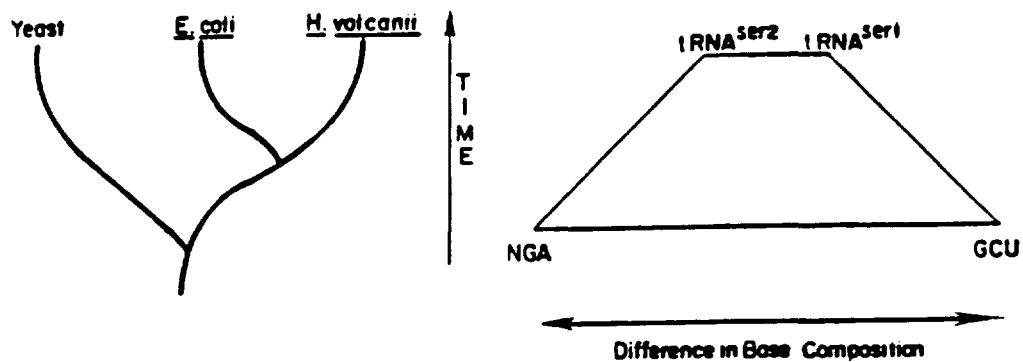
Oblique matches such as these indicate that the evolution of present-day tRNA molecules may have involved the tandem duplication of a shorter nucleotide string. Following this early tandem duplication, mutations have obliterated much of the original pattern, resulting for example in loss of the match at position 35-45 in the methionine initiator.

We have found oblique matches in comparisons among other tRNAs, but always at a low frequency. Of particular interest is the fact that while *H. volcanii* tRNA^{ser1} (GCU) was found to exhibit only two oblique matches when compared with 108 other tRNAs, tRNA^{ser2} (NGA) was found to have 21 oblique matches when compared to the same set of tRNAs. A large number of oblique matches is suggested to indicate a less evolved molecule, *i.e.*, one in which the evidence of early tandem duplication has not been lost by mutation.

There is much evidence that the isoaccepting, serine-encoding tRNAs (anticodes CGA and GCU) form two distinct groups. These include; 1. a low degree of sequence identity between the tRNAs. In fact, in comparisons between forty-three pairs of isoaccepting tRNAs, tRNA^{ser}_{GCU} and tRNA^{ser}_{CGA} showed the least similarity. 2. Differences in relationship with other tRNAs (both in "straight" and oblique matches). 3. Differences in codon usage. It is interesting that tRNA^{ser}_{GCU} is the most frequently used of the two serine tRNA (suggesting that is recognized and used more effectively by the aminoacyl tRNA synthetase). While tRNA^{ser}_{CGA} is the least used. A plausible explanation is that the sequences of these tRNAs have been converging over time, with tRNA^{ser}_{CGA} evolving toward increased efficiency and thus, increased usage. The alternative hypothesis, that tRNA^{ser}_{CGA} is diverging from tRNA^{ser}_{GCU} and thus evolving toward a less efficient molecule, seems less likely.

Other data also suggest an evolutionary convergence of these tRNA molecules. When one compares the sets of tRNA sequences most related to tRNA^{ser1} with the set most related to tRNA^{ser2} an interesting pattern can be seen. Among the *H. volcanii* sequences compared, the two serine encoding tRNAs are essentially related to the same set of sequences. Among *E. coli* sequences, half of each set

is unique to one or the other tRNA^{ser}. Among yeast sequences, each *H. volcanii* tRNA^{ser} has a unique set of most related sequences. A possible interpretation for the differences in relationships between tRNA^{ser1} and tRNA^{ser2} of *H. volcanii* and other tRNAs from *H. volcanii*, *E. coli* and yeast is shown in the figure below.



It is probable that the line of descent leading to yeasts diverged from the line leading to *E. coli* and *H. volcanii* earlier than the divergence of the lines leading to these bacteria. Thus, if one speculates the two tRNAs had different immediate ancestors and have converged to code for serine, the prediction would be that the two *H. volcanii* serine tRNAs would be related to different sets of yeast tRNAs. Further, the list of related *E. coli* tRNAs should overlap considerably and the list of related *H. volcanii* tRNAs should be almost identical. This is precisely what is found.

The results outlined above provide evidence that the NGA and G/ACU assignments for serine are the result of convergent evolution. Convergence of these tRNAs may also explain why the NGA assignments of serine do not fit the hydrophobicity correlations described by Lacey, and Mullins (1983).

Analysis of the tRNA-like terminus of plant RNA viruses.

The tRNA-like ends of many plant RNA virus genomes are receiving much attention especially as some workers consider them to be progenitors to tRNA. The similarities between the viral RNA and tRNAs are striking. Both are recognized by specific enzymes including those which attach amino acids to tRNAs and those which add the universal CCA terminus to tRNAs.

Until recently, it has been thought that the similarity between viral RNA and tRNA is confined to structure and does not extend to base sequence. We have shown that there are significant regions of sequence similarity between tRNAs and some of the tRNA-like viral RNAs. This suggests that the sequences common to the two RNAs may be responsible for enzyme recognition.

Two of the viruses, TMV (tobacco mosaic virus) and TYMV (turnip yellow mosaic virus) contain RNAs whose 3' terminus may be folded into a cloverleaf configuration analogous to the tRNA 2° structure. TMV shows extensive matches with regions corresponding to the T ψ C arm while TYMV has regions corresponding to both the T ψ C and D arms of tRNAs.

Other viruses, which cannot be folded into a cloverleaf, show matches with tRNAs. BBMV (broad bean mottle virus) with the D arm and BMV (brome mosaic virus) with T ψ C and D arms. Interestingly, BMV (which is tyrosylated) contains a ca. 13 base long region in its arm B corresponding to the anticodon region of tyrosine-encoding tRNAs. Other tyrosylating viral RNAs do not show this match with the anticodon arm of tRNA^{tyr}. This raises the question of the significance of the matches. The expected values of the BMV arm B - anticodon matches range from 10⁻³ to 10⁻⁴. Also, only tyrosine-encoding tRNAs participate in such a

match. It seems implausible that tyrosine tRNAs have this unique relationship with tyrosylated viral RNAs by chance.

Convergence seems another unlikely explanation as the BMV RNA is more similar to tRNA^{tyr} from yeast and E. coli than to tRNA^{tyr} from wheat, its host. This suggests common origin as the explanation for the matches. Why then do other tyrosylated viral RNAs not show a match with the anticodon arm of tRNA^{tyr}? Perhaps the sequence has taken on a function in BMV different from that in tRNA. For example, it may be involved in host recognition or interaction with viral proteins. Although this is speculation, it agrees with all known data on the subject.

We are presently pursuing this problem with Dr. Theo Dreher of the Department of Agricultural Chemistry, Oregon State University.

References

Goad, W.B. and Kanehisa, M.I. (1982) *Nucleic Acids Res.* 10:247-263.

Hou, Y-M and Schimmel, P. (1988) *Nature* 333:140-145.

Jukes, T.H. (1966) *Biochem. Biophys Res. Comm.* 24:744-749.

Lacey, J.C., Jr., Hawkins, A.F., Thomas, R.D. and Watkins, C.L., (1988) *Proc. Natl. Acad. Sci. USA* 85. 4996-5000.

Lacey, J.C., Jr. and Mullins, D.W., Jr. (1983) *Orig. Life* 13 3-42.

Staves, M.P., Bloch, D.P. and Lacey, J.C., Jr. (1987) *Z. Naturforsch.* 42C:129-133.

Table I

Summary of Rate Constants
for Formation of Ac-Phe Diester from
Ac-Phe Monoesters of 5'-AMP

Isomeric Identity of Starting Monoester	Isomeric Identity of Second Amino Acid	Rate Const., Min ⁻¹ *		
		<u>1</u>	<u>2</u>	<u>Avg.</u>
D	D	0.03	0.036	0.033
D	L	0.021	0.022	0.022
L	D	0.101	0.098	0.100
L	L	0.07	0.088	0.079

*Duplicate Experiments and Averages

Table II

HPLC Retention Times and % 3' Dipeptidyl
AMP Ester

<u>Dipeptide</u>	Peptide *	Retention Times*		% 3'
		of AMP Ester, Min.		
		<u>2'</u>	<u>3'</u>	
	<u>Retention Time. Min.</u>			<u>Ester</u>
Ac-L-Phe-L-Phe	34	25.5	32.1	61.7
Ac-L-Phe-D-Phe	58	44.0	65.0	49.8
Ac-D-Phe-D-Phe	35	36.0	41.0	48.5
Ac-D-Phe-L-Phe	60	46.0	67.0	61.6

*Phenyl μ Bondapak column 3.9 mm x 30 mm 5 μ particles. Elution solvent was 30% CH₃OH, 0.05 phosphoric acid, pH 2 at 1.0 ml/min. Monitored at 259nm.

	ALA1	ALA2	ARG1	ARG2	ASN1	ASP1	CYS1	GLN1
ALA1	---	3.8E-33	1.5E-10	6.5E-10*	7.2E-4	1.6E-9	2.6E-4	0
ALA2	3.8E-33	---	8.0E-11	3.9E-10*	6.9E-4	7.3E-9	2.2E-4	8.6E-4
ARG1	1.5E-10	8.9E-11	---	3.3E-43	3.9E-9	2.0E-9*	5.4E-4	7.3E-8
ARG2	6.5E-10*	3.9E-10*	3.3E-43	---	4.5E-4	2.7E-10*	5.3E-4	6.9E-8
ASN1	7.2E-4	6.9E-4	3.9E-9	4.5E-4	---	9.4E-16	1.5E-7	4.5E-4
ASP1	1.6E-9	7.3E-9	2.0E-9*	2.7E-10*	9.4E-16	---	5.8E-4	1.9E-4
CYS1	2.6E-4	2.2E-4	5.4E-4	5.3E-4	1.5E-7	5.8E-4	---	5.0E-9
GLN1	0	8.6E-4	7.3E-8	6.9E-8	4.5E-4	1.9E-4	5.0E-9	---
GLN2	0	8.5E-4	2.2E-5	2.2E-5	1.0E-6	6.4E-4	5.1E-4	1.7E-32
GLU1	5.1E-4	3.6E-4	3.2E-5	3.0E-5	4.7E-4	7.7E-5	6.2E-4	2.9E-5
GLU2	4.5E-4	2.3E-4	3.0E-5	2.9E-5	4.5E-4	6.8E-5	6.0E-4	2.7E-5
GLY1	2.6E-4	2.2E-4	2.2E-8*	5.2E-4	4.9E-11*	9.1E-11*	4.9E-4	1.4E-4
GLY3	3.8E-6	3.0E-6	5.4E-4	5.2E-4	4.6E-4	1.1E-6	5.0E-4	5.6E-4
GLY5	1.3E-7	7.8E-11*	1.3E-11*	1.9E-10*	3.3E-9*	1.4E-12*	0	1.1E-4
HIS1	1.5E-12	3.5E-10	3.5E-13*	4.7E-12*	3.5E-11	2.4E-10*	4.5E-4	1.3E-9
ILE1	4.5E-10*	3.0E-10*	9.8E-12*	3.8E-12*	1.2E-6	3.5E-10*	0	0
ILE2	1.0E-4	9.5E-9*	4.3E-10*	1.8E-10*	6.6E-13	1.5E-11*	0	0
LEU1	0	0	0	0	1.7E-5	1.1E-7	0	0
LEU2	2.8E-4	0	2.2E-4	2.1E-4	5.3E-4	7.3E-5	0	0
LEU3	2.0E-4	1.5E-4	0	0	4.3E-7*	1.5E-7	1.8E-4	0
LYS1	1.3E-13	3.3E-12*	6.4E-15	1.2E-14	2.6E-13	4.0E-11	4.2E-5	9.4E-7
METE	4.5E-10*	4.8E-11*	5.5E-14*	1.5E-14*	7.4E-12*	2.9E-14	4.3E-4	8.6E-7
MET1	0	0	2.0E-8*	3.2E-5	8.5E-5	3.5E-4	2.2E-4	7.1E-4
PHE1	1.5E-9*	9.3E-10*	2.1E-9*	2.8E-8*	2.2E-14	3.7E-10*	2.5E-9	1.8E-7
SER1	3.1E-5	3.5E-4	7.0E-6	6.8E-6	5.0E-4	7.8E-5	6.0E-4	2.6E-7
SER2	0	0	3.8E-4	3.8E-4	0	4.0E-5	0	4.9E-12*
SER3	9.0E-4	0	3.8E-5	3.7E-5	0	2.9E-4	0	6.3E-4
THR1	1.3E-13	2.6E-14	4.4E-13*	6.0E-12*	4.7E-6	2.6E-4	0	1.4E-7*
TRP1	1.3E-6	2.4E-5	1.3E-9	2.1E-8	8.4E-10*	3.9E-9	0	6.4E-4
TYR1	0	8.6E-4	1.1E-7	1.1E-7	4.0E-8	6.1E-4	1.8E-7	7.6E-11
VAL1	2.0E-9	9.0E-10	3.9E-19	6.4E-20	4.8E-12*	4.3E-14	5.5E-4	5.9E-4
VAL2	1.2E-18	1.2E-19	4.7E-12	1.9E-11	3.7E-5	2.2E-10	2.1E-4	1.2E-4
VAL3	8.7E-11	2.4E-10	1.0E-11*	2.7E-12*	1.4E-10*	1.4E-12*	4.5E-4	4.6E-4

Table 3. A matrix showing the calculated E values for comparisons among 35 *E. coli* tRNAs. An asterisk indicates the value is derived from E values for two matching regions.

	GLN2	GLU1	GLU2	GLY1	GLY3	GLY5	HIS1	ILE1
ALA1	0	5.1E-4	4.5E-4	2.6E-4	3.8E-6	1.3E-7	1.5E-12	4.5E-10*
ALA2	8.5E-4	3.6E-4	2.3E-4	2.2E-4	3.0E-6	7.8E-11*	3.5E-10	3.0E-10*
ARG1	2.2E-5	3.2E-5	3.0E-5	2.2E-8*	5.4E-4	1.3E-11*	3.5E-13*	9.8E-12*
ARG2	2.2E-5	3.0E-5	2.9E-5	5.2E-4	5.2E-4	2.0E-10*	4.7E-12*	3.8E-12*
ASN1	1.0E-6	4.7E-4	4.5E-4	4.9E-11*	4.6E-4	3.3E-9*	3.5E-11	1.2E-6
ASP1	6.4E-4	7.7E-5	6.8E-5	9.1E-11*	1.1E-6	1.4E-12*	2.4E-10*	3.5E-10*
CYS1	5.1E-4	6.2E-4	6.0E-4	4.9E-4	5.0E-4	0	4.5E-4	0
GLN1	1.7E-32	2.9E-5	2.7E-5	1.4E-4	5.6E-4	1.1E-4	1.3E-9	0
GLN2	---	2.7E-5	2.6E-5	5.1E-4	5.5E-4	0	1.8E-5	5.2E-4
GLU1	2.7E-5	---	4.3E-41	5.1E-5	1.9E-5	0	1.2E-6	0
GLU2	2.6E-5	4.3E-41	---	4.9E-5	5.4E-5	0	1.1E-6	0
GLY1	5.1E-4	5.1E-5	4.9E-5	---	8.7E-14	2.2E-15	2.4E-6	0
GLY3	5.5E-4	1.9E-5	5.4E-5	8.7E-14	---	2.1E-8	3.9E-5	0
GLY5	0	0	0	2.2E-15	2.1E-8	---	3.6E-14*	2.9E-12
HIS1	1.8E-5	1.2E-6	1.1E-6	2.4E-6	3.9E-5	3.6E-14*	---	2.2E-10*
ILE1	5.2E-4	0	0	0	0	2.9E-12	2.2E-10*	---
ILE2	1.5E-5	0	0	6.9E-4	4.9E-4	1.5E-11*	2.6E-10*	1.1E-14
LEU1	0	6.9E-4	6.4E-4	6.0E-8*	4.8E-5	3.9E-10*	5.6E-7*	2.4E-5
LEU2	0	8.6E-5	7.8E-5	6.1E-4	2.0E-4	0	6.2E-7*	1.7E-5
LEU3	0	8.3E-4	7.9E-4	5.8E-4	3.6E-7	2.4E-4	5.5E-4	2.7E-5
LYS1	1.6E-5	1.5E-5	1.4E-5	4.2E-4	4.5E-5	5.8E-15*	3.9E-15	4.6E-11*
METE	1.7E-5	4.5E-6	4.5E-6	1.2E-5	1.1E-8*	8.1E-16	2.1E-16	3.3E-13
MET1	6.9E-4	0	0	0	4.5E-4	0	4.1E-6	7.5E-6
PHE1	1.7E-5	7.1E-4	6.7E-4	8.2E-9*	2.0E-5	6.3E-9	1.4E-14	2.6E-5
SER1	2.5E-5	2.0E-7	1.8E-7	1.5E-6	4.5E-5	0	1.2E-6	1.9E-6
SER2	2.8E-12	4.3E-4	4.1E-4	5.3E-5	0	5.3E-5	2.9E-4	0
SER3	6.2E-4	6.7E-6	6.3E-6	1.7E-5	1.9E-4	0	1.2E-5	6.5E-5
THR1	1.5E-7*	3.1E-4	2.9E-4	0	0	1.9E-16	1.2E-14*	3.8E-18
TRP1	1.9E-4	0	0	1.1E-14	1.7E-4	6.5E-11	1.3E-6	1.1E-4
TYR1	1.2E-9	3.0E-5	2.9E-5	3.0E-6	5.4E-4	0	1.6E-5	0
VAL1	4.7E-5	9.5E-5	8.5E-5	5.4E-4	5.7E-4	1.3E-17	7.6E-13*	1.1E-18
VAL2	8.1E-4	1.9E-8	1.6E-8	1.5E-4	5.9E-5	1.3E-12	1.2E-11*	4.3E-11
VAL3	3.5E-5	4.8E-4	4.6E-4	4.5E-4	4.6E-4	9.5E-19	1.1E-15*	5.5E-16

	ILE2	LEU1	LEU2	LEU3	LYS1	METE	METI	PHE1
ALA1	1.0E-4	0	2.8E-4	2.0E-4	1.3E-13	4.5E-10*	0	1.5E-9*
ALA2	3.3E-12*	4.8E-11*	0	9.3E-10*	9.5E-9*	0	0	1.5E-4
ARG1	4.3E-10*	0	2.2E-4	0	6.4E-15	5.5E-14*	2.0E-8*	2.1E-9*
ARG2	1.8E-10*	0	2.1E-4	0	1.2E-14	1.5E-14*	3.2E-5	2.8E-8*
ASN1	6.6E-13	1.7E-5	5.3E-4	4.3E-7*	2.6E-13	7.4E-12*	8.5E-5	2.2E-14
ASP1	1.5E-11*	1.1E-7	7.3E-5	1.5E-7	4.0E-11*	2.9E-14	3.5E-4	3.7E-10*
CYS1	0	0	0	1.8E-4	4.2E-5	4.3E-4	2.2E-4	2.5E-9
GLN1	0	0?	0	0	9.4E-7	8.6E-7	7.1E-4	1.8E-7*
GLN2	1.5E-5	0	0	0	1.6E-5	1.7E-5	6.9E-4	1.7E-5
GLU1	0	6.9E-4	8.6E-5	8.3E-4	1.5E-5	4.5E-6	0	7.1E-4
GLU2	0	6.4E-4	7.8E-5	7.9E-4	1.4E-5	4.5E-6	0	3.7E-4
GLY1	6.9E-4	6.0E-8*	6.1E-4	5.8E-4	4.2E-4	1.2E-5	0	8.2E-9*
GLY3	4.9E-4	4.8E-5	2.0E-4	3.6E-7	4.5E-5	1.1E-8*	4.5E-4	2.0E-5
GLY5	1.5E-11*	3.9E-10*	0	2.4E-4	5.8E-15*	8.1E-16	0	6.3E-9
HIS1	2.6E-10*	5.6E-7*	6.2E-7*	5.5E-4	3.9E-15	2.1E-16	4.1E-6	1.4E-14
ILE1	1.1E-14	2.4E-5	1.7E-5	2.7E-5	4.6E-11*	3.3E-13	7.5E-6	2.6E-5
ILE2	---	2.1E-5	2.8E-5	3.3E-10	3.4E-19	1.7E-14	8.6E-6	1.1E-13
LEU1	2.1E-5	---	6.1E-19	1.3E-17	0	4.8E-10*	6.6E-4	0
LEU2	2.8E-5	6.1E-19	---	1.7E-12*	0	0	1.6E-5	0
LEU3	3.3E-10	1.3E-17	1.7E-12*	---	3.9E-4	5.2E-4	9.6E-5	1.8E-14
LYS1	3.4E-19	0	0	3.9E-4	---	9.6E-6	3.8E-6	7.6E-13*
METE	1.7E-14	4.8E-10*	0	5.2E-4	9.6E-6	---	4.9E-4	5.3E-11*
METI	8.6E-6	6.6E-4	1.6E-5	9.6E-5	3.8E-6	4.9E-4	---	2.6E-4
PHE1	1.1E-13	0	0	1.8E-14	7.6E-13*	5.3E-11*	2.6E-4	---
SER1	0	9.3E-5	3.7E-5	0	1.6E-5	4.5E-6	0	7.2E-4
SER2	0	0	0	0	1.6E-6	1.4E-5	0	0
SER3	6.4E-4	5.6E-5	4.1E-7	8.4E-6	3.9E-4	1.4E-4	4.1E-5	0
THR1	1.2E-6	0	5.7E-5	0?	2.1E-16*	6.9E-13	1.8E-5	1.5E-8
TRP1	6.2E-5	1.8E-5	1.7E-8*	5.8E-4	5.9E-8	1.2E-7	0	6.8E-6
TYR1	0	0	1.2E-4	0	1.4E-7	4.3E-6	1.1E-6	5.6E-4
VAL1	1.5E-15*	5.7E-9	2.8E-4	6.8E-4	2.8E-14*	3.0E-17	2.9E-4	4.3E-11*
VAL2	6.9E-6	4.4E-8	0	0	3.2E-16	5.7E-16	2.6E-4	7.9E-10*
VAL3	6.6E-16*	2.8E-9	5.6E-4	5.1E-4	2.8E-14*	4.2E-16	1.7E-5	1.4E-13*

	SER1	SER2	SER3	THR1	TRP1	TYR1	VAL1	VAL2	VAL3
ALA1	3.1E-5	0	9.0E-4	1.3E-13	1.3E-6	1.0E-3	2.0E-9	1.2E-18	8.7E-11
ALA2	3.5E-4	0	0	2.6E-14	2.4E-5	8.6E-4	9.0E-10	1.2E-19	2.4E-10
ARG1	7.0E-6	3.8E-4	3.8E-5	4.4E-13*	1.3E-9	1.1E-7	3.9E-19	4.7E-12	1.0E-11
ARG2	6.8E-6	3.8E-4	3.7E-5	6.0E-12*	2.1E-8	1.1E-7	6.4E-20	1.9E-11	2.7E-12*
ASN1	5.0E-4	0	0	4.7E-6	8.4E-10*	4.0E-8	4.8E-12*	3.7E-5	1.4E-10*
ASP1	7.8E-5	4.0E-5	2.9E-4	2.6E-4	3.9E-9	6.1E-4	4.3E-14	2.2E-10	2.8E-14
CYS1	6.0E-4	0	0	0	0	1.8E-7	5.5E-4	2.1E-4	4.5E-4
GLN1	2.6E-7	4.9E-12*	6.3E-4	1.4E-7*	6.4E-4	7.6E-11	5.9E-4	1.2E-4	4.6E-4
GLN2	2.5E-5	2.8E-12	6.1E-4	1.5E-7*	1.9E-4	1.2E-9	4.7E-5	8.1E-4	3.5E-5
GLU1	2.0E-7	4.3E-4	6.7E-6	3.1E-4	0	3.0E-5	9.5E-5	1.9E-8	4.8E-4
GLU2	1.8E-7	4.1E-4	6.3E-6	2.9E-4	0	2.9E-5	8.5E-5	1.6E-8	4.6E-4
GLY1	1.5E-6	5.3E-5	1.7E-5	0	1.1E-14	3.0E-6	5.4E-4	1.5E-4	4.5E-4
GLY3	4.5E-5	0	1.9E-4	0	1.8E-4	5.4E-4	5.7E-4	5.9E-5	4.6E-4
GLY5	0	5.3E-5	0	1.9E-16	6.5E-11	0	1.3E-17	1.3E-12	9.5E-19
HIS1	1.2E-6	2.9E-4	1.2E-5	1.2E-14*	1.3E-6	1.6E-5	7.6E-13*	1.2E-11*	1.1E-15*
ILE1	1.9E-6	0	6.5E-5	3.8E-18	1.1E-4	0	1.1E-18	4.3E-11	5.5E-16
ILE2	0	0	6.4E-4	1.2E-6	6.2E-5	0	1.5E-15*	6.9E-6	6.6E-16*
LEU1	9.3E-5	0	5.6E-5	0	1.8E-5	0	5.7E-9	4.4E-8	2.8E-9
LEU2	3.7E-5	0	4.1E-7	5.7E-5	1.7E-8*	1.2E-4	2.8E-4	0	5.6E-4
LEU3	0	0	8.4E-6	0	5.8E-4	0	6.8E-4	0	5.1E-4
LYS1	1.6E-5	1.6E-6	3.9E-4	2.1E-16*	5.9E-8	1.4E-7	2.8E-14*	3.2E-16	2.8E-14*
METE	4.5E-6	1.4E-5	1.4E-4	6.9E-13	1.2E-7	4.3E-6	3.0E-17	5.7E-16	4.2E-16
METI	0	0	4.1E-5	1.8E-5	0	1.1E-6	3.0E-4	2.6E-4	1.7E-5
PHE1	7.2E-4	0	0	1.5E-8	6.8E-6	5.6E-4	4.3E-11*	7.9E-10*	1.4E-13*
SER1	—	1.6E-11*	1.4E-13*4.0E-9	9.2E-5	1.5E-10	2.5E-4	4.9E-5	5.3E-4	
SER2	1.6E-11*	—	2.9E-11	2.0E-5	1.1E-13	4.9E-7*	0	0	0
SER3	1.4E-13*	2.9E-11	—	0	4.9E-12	7.9E-4	7.9E-4	0	0
THR1	4.0E-9	2.0E-5	0	—	3.6E-11	2.8E-4	1.1E-12	1.0E-11	7.1E-14
TRP1	9.2E-5	1.1E-13	4.9E-12	3.6E-11	—	0	1.8E-11*	7.8E-6	9.7E-11*
TYR1	1.5E-10	4.9E-7*	7.9E-4	2.8E-4	0	—	5.7E-4	8.5E-4	4.4E-4
VAL1	2.5E-4	0	7.9E-4	1.1E-12	1.8E-11*	5.7E-4	—	4.2E-15	4.3E-32
VAL2	4.9E-5	0	0	1.0E-11	7.8E-6	8.5E-4	4.3E-15	—	1.3E-14
VAL3	5.3E-4	0	0	7.1E-14	9.7E-11*	4.4E-4	4.3E-32	1.3E-14	—

ORIGINAL PAGE IS
OF POOR QUALITY

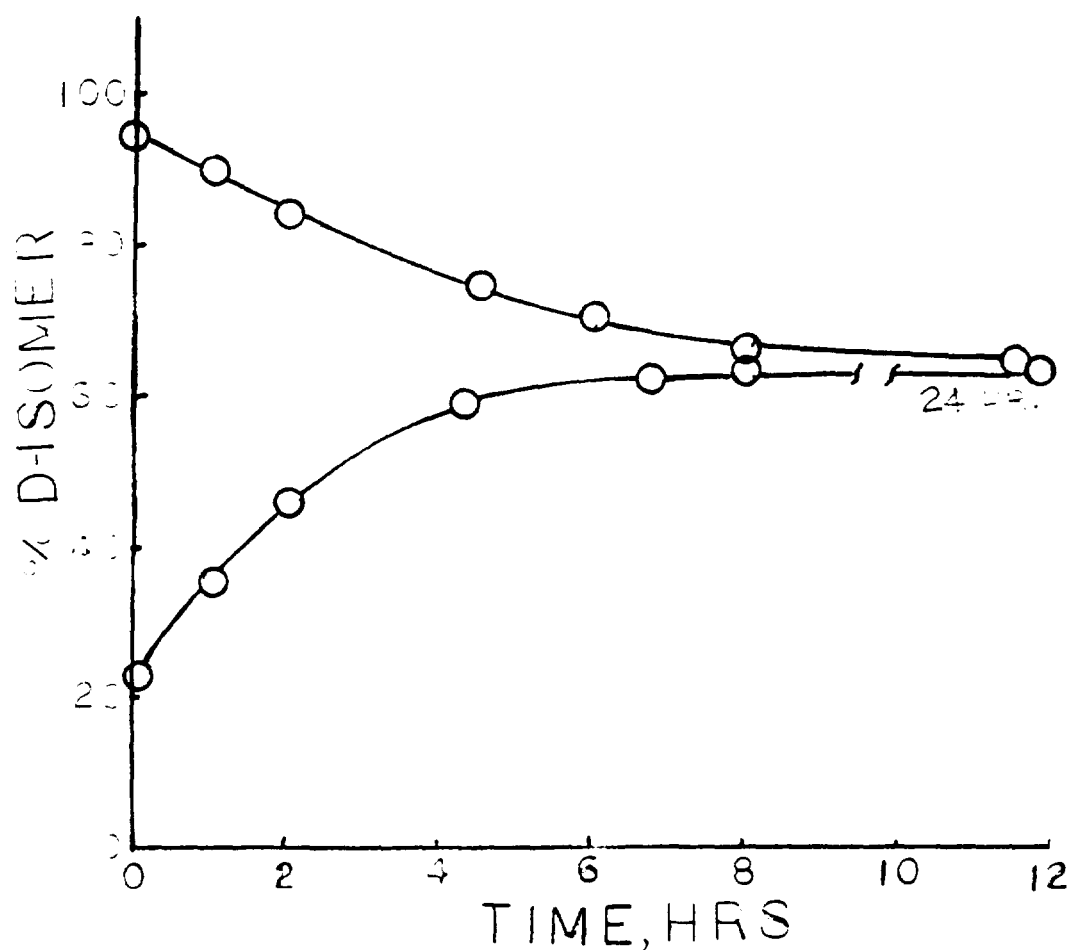


Fig. 1. Percent D-isomer in esterifications of 5' AMP run at various times after the preparation of the activated form, Ac-Phe imidazolide. The percent D-and L-forms were determined by HPLC separations of samples removed from the esterification reaction at 5 min. Reactions were at 0 C, pH 7.4. 5'-AMP concentration was 0.009 M and Ac-Phe-imidazolide was 0.09 M in the esterifications. The percent D-isomer is plateauing at about 65%. Because we know that the Ac-Phe-imidazolide becomes racemic with time (in this case after about 6 Hr.) that means the Ac-D Phe is esterifying the AMP $65/35=1.86$ times as fast as the Ac-L-Phe.

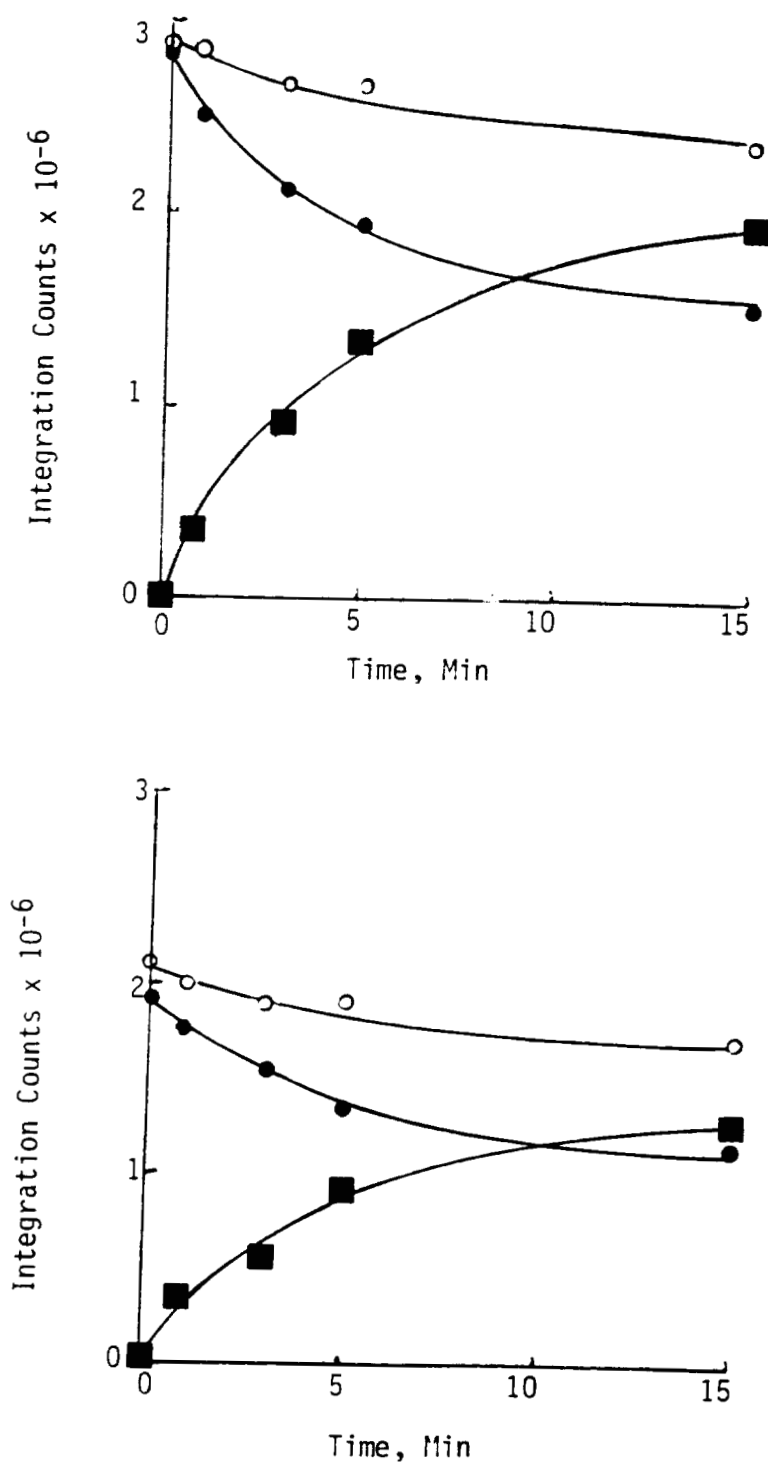


Fig. 2 Plots of HPLC integration counts for the appearance of total diester (■) and for disappearance of Ac-D-PheAMP monoester (○) and Ac-L-PheAMP monoester (●) when reacting the mixed monoester with a 10-fold excess of Ac-D-Phe Imidazolidine (A) or Ac-L-Phe-Imidazolidine (B) at 0°C and pH 7.4. Initial monoester conc, 10 μ moles/ml. This data is from one of the runs in Table I.

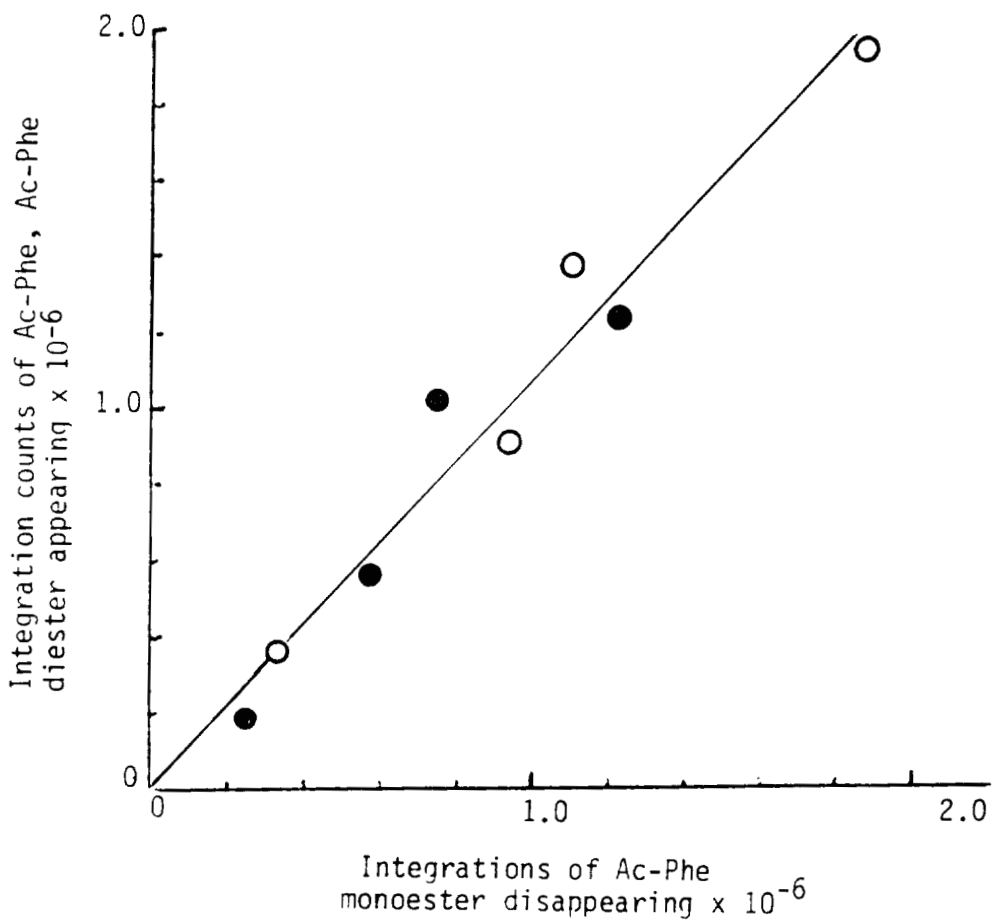


Fig.³ .Data from Fig. 2 plotted to show the equality of disappearance of monoester and appearance of diester. This plot shows that the diester is arising solely from monoester and not from free AMP which was previously removed by SEP-PAK separation.

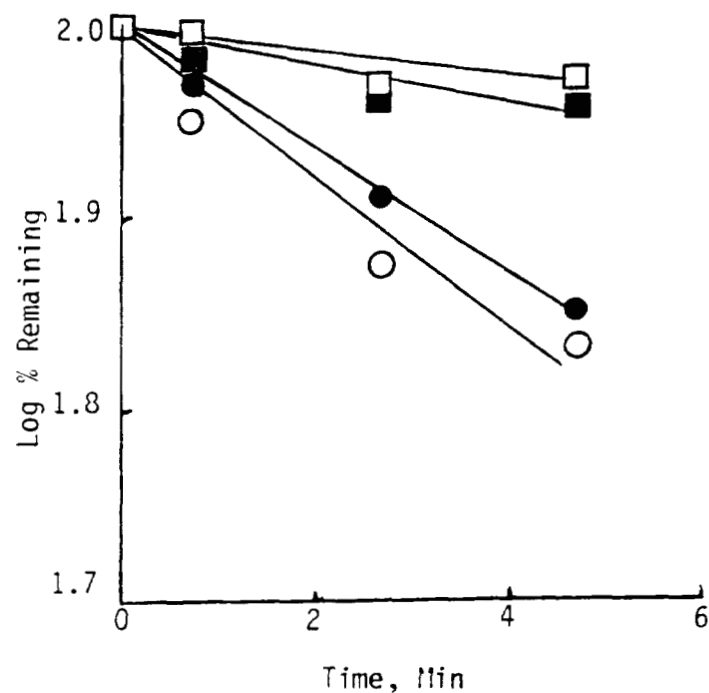


Fig. 4. Pseudo first order plot of the data in Fig. 2 for the disappearance of AC-D-Phe-AMP monoester (□, ■) and for the disappearance of Ac-L-PheAMP monoester (○, ●) when treated with Ac-D-Phe imidazolidine (open points) or Ac-L-Phe-imidazolidine (closed points) to form the Ac-Phe diester. Initial imidazolidine conc. 100 μ moles/ml, monoester conc. 10 μ moles/ml at pH 7.4 and 0°C. Reactions contained about 15% dimethylformamide which was not necessary but was carried over from the imidazolidine synthesis.

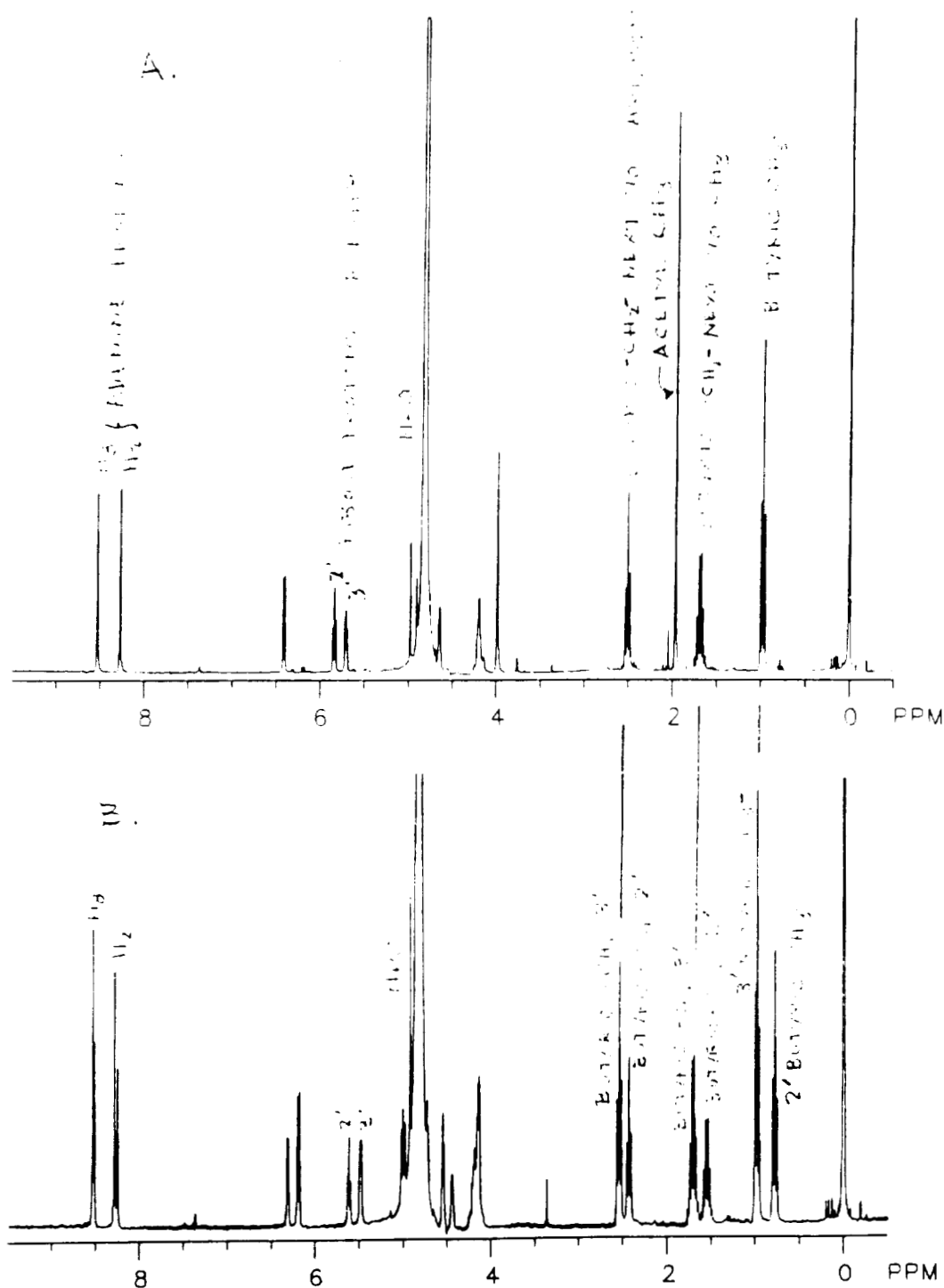
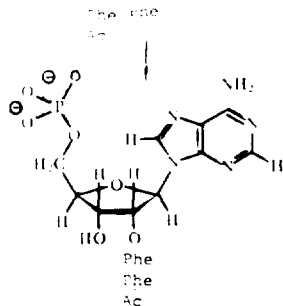
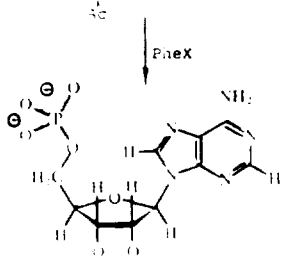
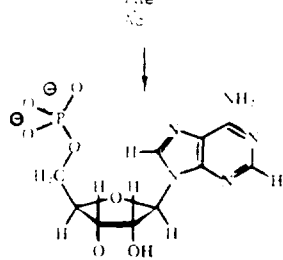
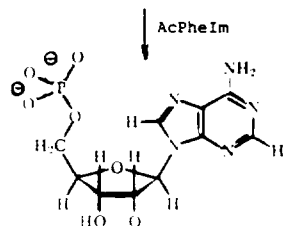
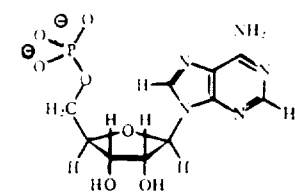


Fig. 5 PMR spectra of A. Acetylglycine and butyric acid ester of 5'-AMP and B. butyric acid monoester of 5'-AMP. Spectra were run on Nicolet 300 MHz spectrometer in the pulse mode. Samples were 0.004 M in 0.05 M phosphate at pH 5. In the monoester spectrum (B), the butyric acid distributes between the 2' (40%) and 3' (60%) giving rise to two sets of peaks for each kind of butyric proton, the larger set in each case is the 3' ester. It can be seen in the upper spectrum that the butyric acid signals correspond to the 3' ester peaks in the lower spectrum. These data show conclusively that the butyric acid which was esterified to the AMP first (see A.) is now in the 3' position and the Ac-Gly, added second, is in the 2' position.

Fig. 6 FLOWSHEET FOR THE PREFERENTIAL CATALYSIS OF THE SYNTHESIS OF
L-AMINO ACID PEPTIDES ON THE RIBOSE OF 5'-AMP
USING THE SYNTHESIS OF TRIPHENYLALANINE AS AN
EXAMPLE

ORIGINAL PAGE IS
OF POOR QUALITY

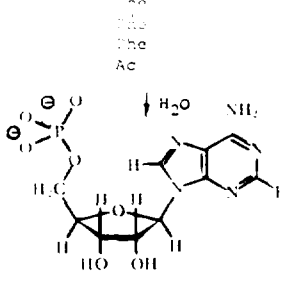
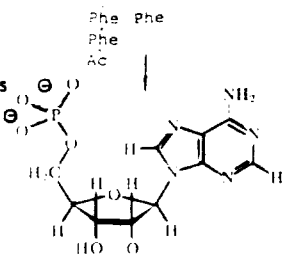
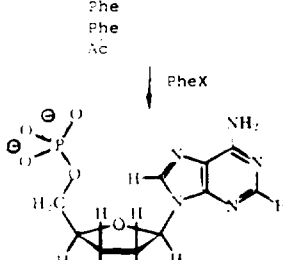
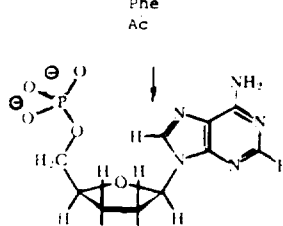
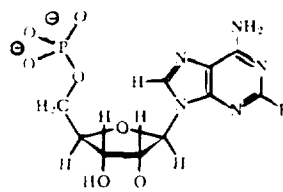


Esterification of the 2' position of 5'-AMP yielding a monoester of AcPhe. Data suggest esterification is exclusively at the 2' position.

Redistribution of the Ac-Phe to the 3' position leaving the 2' position open for diester formation. If the Ac-Phe is L-isomer, 67% is in the 3' position, if D-isomer 50% is in the 3' position

Reaction of the 2' position with Phe-x. The 2' position has been shown to be almost exclusive position of attack to form diester. Consequently a preexisting L-mono ester reacts 3 times as rapidly as D to yield diester. Experimentally verified.

Formation of Ac-Phe-Phe by nucleophilic attack of L-amino acid in 2' position on the carboxyl of the Ac-Phe in the 3' position, yielding Ac-Phe-Phe in the 2' position. Data show the peptide can form.



+
AcPhePhePhe

Distribution of the peptide to the 3' position. If the newly added amino acid is L, 67% of the peptide will be in the 3' position, if D, 50%

Reaction with another activated amino acid esterifying the 2' position with an amino acid having a free NH₂ (α amino), again preexisting L-amino acid peptide would react three times as fast as preexisting D-peptide.

Formation of another peptide bond yielding a blocked tripeptide in the 2' position. The cycle of redistribution to 3', aminoacylation of 2' and peptide bond formation could continue indefinitely, yielding large peptides and finally:

Hydrolysis of the peptide regenerating the 5'-AMP unchanged and releasing the peptide.

While the ability of AMP to catalyze the synthesis of peptides is possible simply because the diester can be formed and adjacent reactive species can react with each other, the ability to preferentially synthesize L-based peptides depends on two factors; (1) all esterifications occur at the 2' position and (2) all L-amino acids (and Glycine) distribute preferentially to the 3' position (~67%). Consequently with L-amino acids (or peptides) 67% of the reactive 2' position is available for formation of the diester. With D-amino acids (or peptides) the percent of the 2' position available is not constant but varies inversely as the hydrophobicity of the amino acid, Trp, for example only having 40% of the 2' position available. All steps have been experimentally verified, but whether there is any difference in

rates of peptide formation depending on the combination of optical isomers is not yet known ie does the combination LL react faster than combination DL etc? There may well be selectivity at the point of peptide bond formation also. This ability of 5'-AMP (and perhaps other ribonucleotides) to catalyze the synthesis of peptides may well be the most primitive and fundamental relationship between the protein and nucleic acid system. Furthermore the origin and evolution of the process of protein synthesis may simply be an elaboration of this ability. Built into the relationship is not only the basis for synthesis of the peptide bond but also selectivity for L-isomers and the basis for growth from amino to carboxy ie direction of protein synthesis.

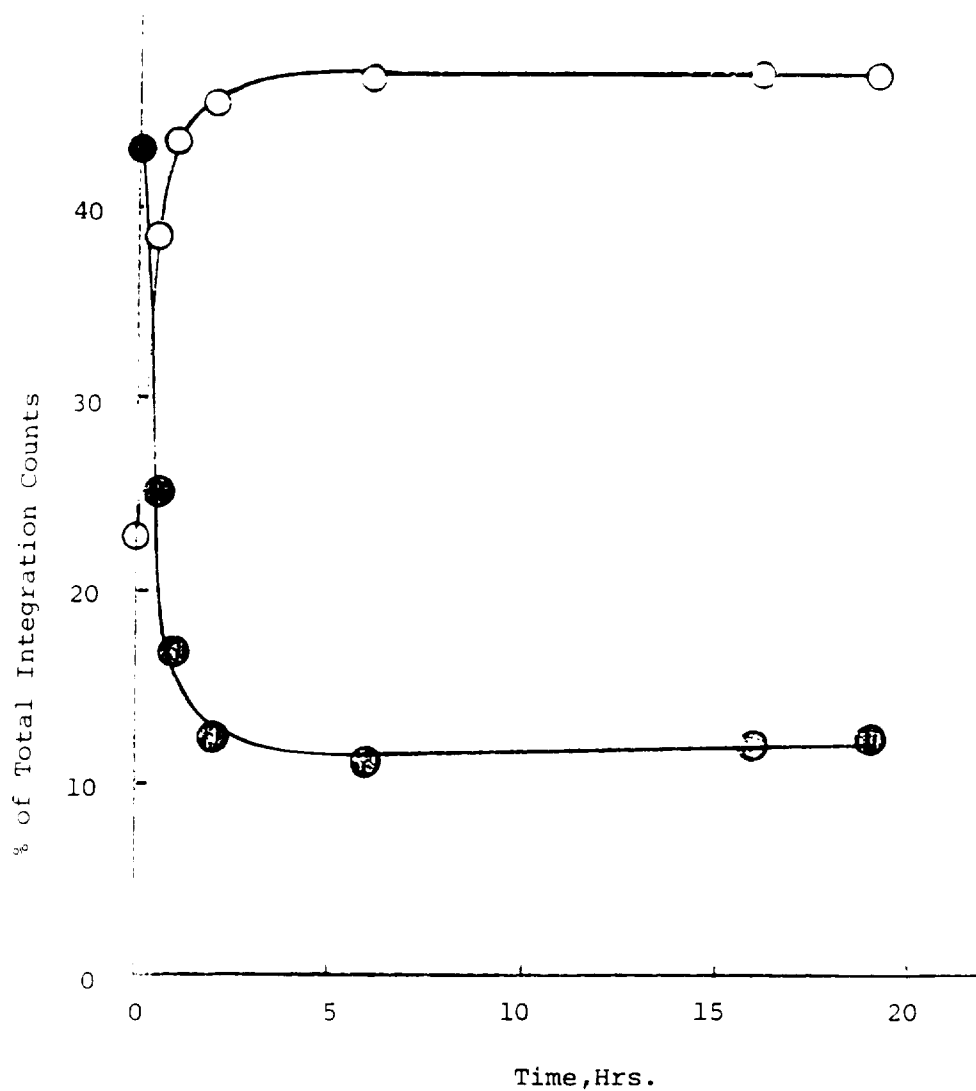


Fig. 7. The disappearance of HPLC integration counts(●) identified with Ac-L-Phe-AMP monoester(3.8 min. peak) with time and the appearance of integration counts(○) identified with the putative 3'Ac-Phe-2'tBOC Phe-AMP diester(18 min. peak) with time. Reaction conditions and HPLC conditions are described in the text.

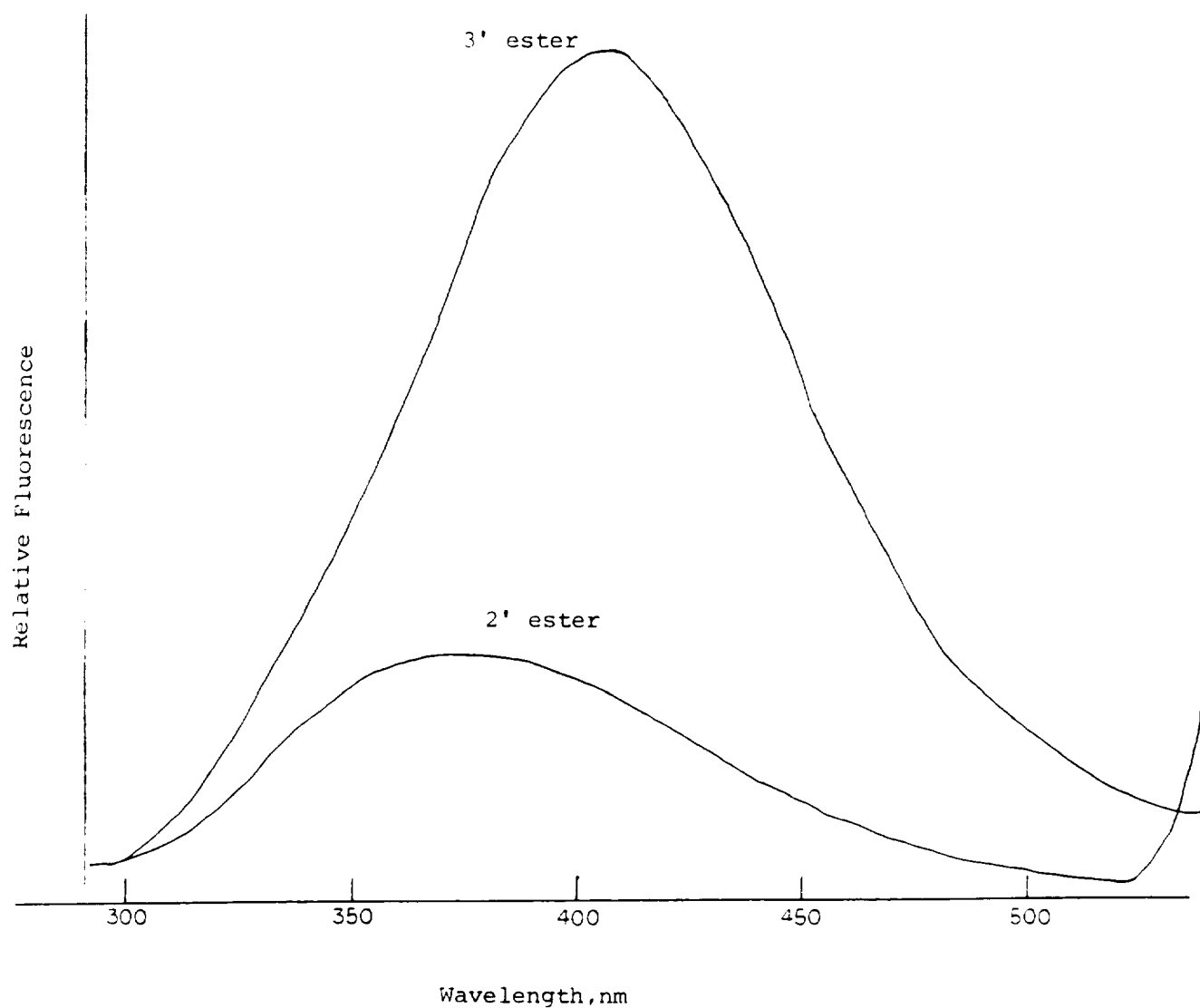


Fig.8. Relative fluorescent spectra of the 3'-Ac-D-Trp-AMP and 2'-Ac-D-Trp-AMP monoesters when exciting at 275 nm. Samples were identical concentrations (about 10^{-6} M) and were at pH 2 in 20% CH_3OH ; 0.05 M phosphoric acid. Spectra were recorded on a Perkin-Elmer 650-40 Fluorescence Spectrometer.

Hydrolysis of Phenyl-3-Propionic Acid - AMP Anhydride

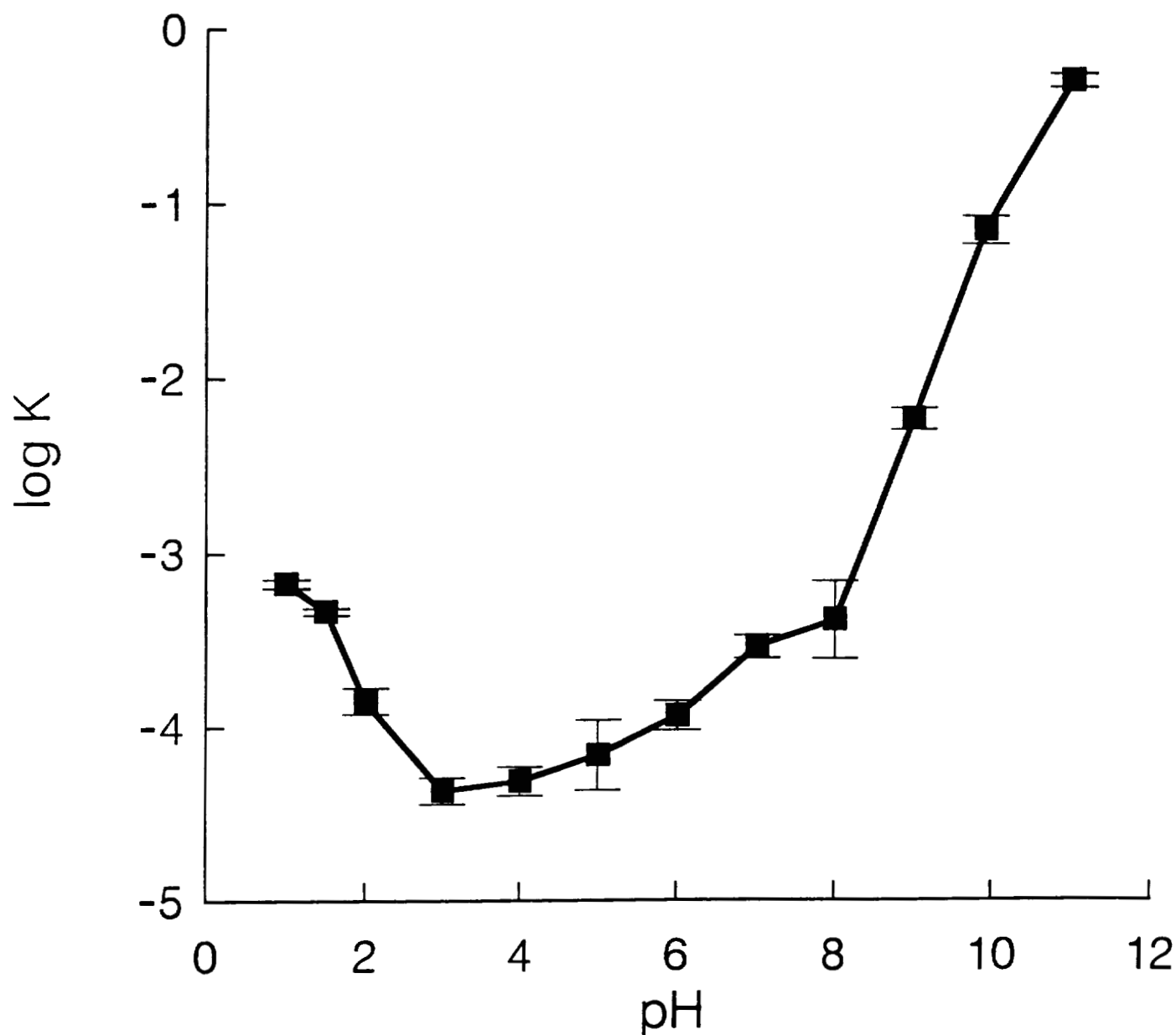


Fig. 9. Log $k(\text{min.}^{-1})$ of the hydrolysis constants(pseudo first order) plotted vs. the pH at 25°C . Each point represents the log k for the rate constant run at three different buffer concentrations (0.0125, 0.025, and 0.05 M) and so there is very little salt or specific ion effect on the rate of hydrolysis.

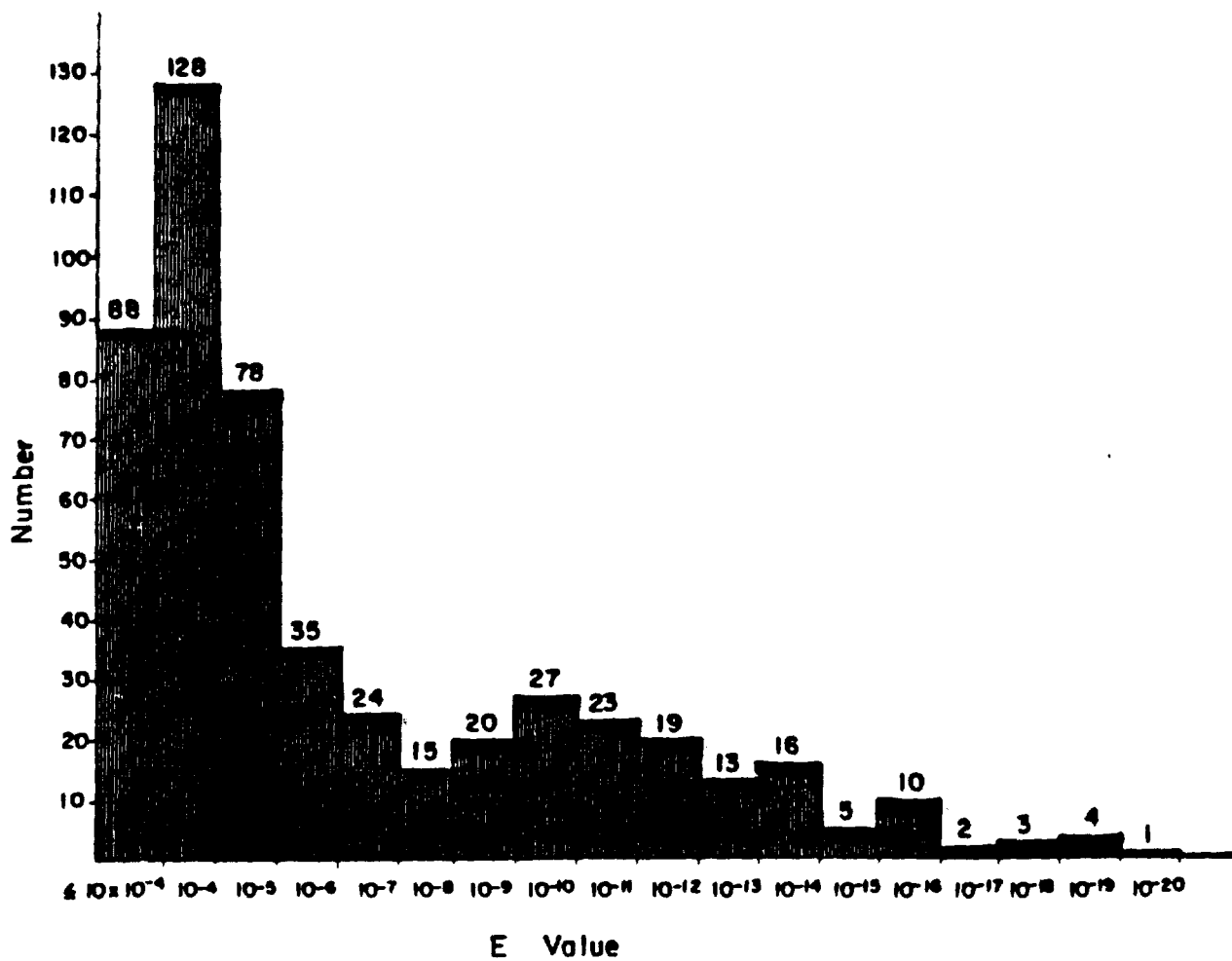


Figure 10. A histogram showing the distribution of E values calculated from comparisons among 35 E. coli tRNAs. The procedure for calculating E is described in the text. Note the peaks at $10^{-3} > x > 10^{-4}$ and $10^{-9} > x > 10^{-11}$.

These are due in large part to conservation of bases in the D and TΨC loops.

RESEARCH PROPOSAL SUBMITTED TO:

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

by

University of Alabama School of Medicine
University Station
Birmingham, Alabama 35294

TITLE OF PROPOSED PROJECT:

THE CHEMICAL BASIS FOR THE ORIGIN OF THE
GENETIC CODE AND THE PROCESS OF
PROTEIN SYNTHESIS

PRINCIPAL INVESTIGATOR:

NAME: James C. Lacey, Jr., Ph.D.

TITLE: Associate Professor of Biochemistry

SOCIAL SECURITY NO:




RENEWAL REQUEST:

PROPOSED STARTING DATE: December 1, 1987

PROPOSED DURATION: Three years

SUBMISSION DATE: September 10, 1987


James C. Lacey, Jr., Ph.D.
Associate Professor of Biochemistry
Department of Biochemistry


Kenneth M. Pruitt, Ph.D.
Asst. Vice President - Research

Abstract

For many years now we have attempted to elucidate the molecular basis for the origin of the genetic code and the process of protein synthesis. Although we are not at a point of saying just what the molecular basis was, several findings have clarified certain aspects of the problem. For example, correlations of properties of amino acids and their anticodonic nucleotides, while not answering the question of mechanism, make it obvious that the genetic code is not a random matrix but has a definite logic to it. This is a first level answer, which suggests the underlying logic has a molecular basis which can be discerned. Determination of binding constants between amino acids and nucleotides has revealed definite selectivities which tend to be anticodonic. Reaction rates between amino acids and nucleotides have also been shown to be selective with an anticodonic bias i.e. some amino acids tend to react faster with their anticodons.

Numbers of hydrophobic intramolecular interactions have been observed showing that hydrophobicity is a definite organizing principle. This kind of interaction has also been shown to have some chiral selectivity offering at least a partial explanation of the origin of the use of L-amino acids in addition to explaining the origin of several aspects of contemporary protein synthesis.

The proposed program is designed principally to finish up several projects already begun.

These are:

1. Completion of a study of the 2'-3' distribution of D and L amino acids and carboxylic acids esterified to mononucleotides. These studies relate to the origin of several factors regarding the 3' end of tRNA and offer one explanation for the use of L-amino acids in protein synthesis.
2. Structural studies of D- and L-aminoacyl-nucleotide esters, principally 5'-AMP, as a model of the end of the tRNA. These studies will use NMR, and X-ray diffraction.
3. Oligopeptide-oligonucleotide interactions. These studies will attempt to determine the binding constants between a tetrapeptide and its codonic and anticodonic DNA strands. NMR and ORD will be the principal tools. These studies will attempt to determine whether the amino acid - codon or anticodon binding is stronger and whether the 3/1 coding ratio is a fundamental thing or not.
4. Starting with aminoacyl-AMP esters we will study peptide bond formation using D and L amino acids at various concentrations and determine whether a poly U template will enhance peptide bond formation. This study attempts to establish a model of primitive protein synthesis.
5. We will continue to study tRNA evolution and investigate 5s rRNA as a possible primitive universal translator molecule and source of tRNAs.

Summary of Previous Work

Because this project has continued for many years now, a detailed summary of the work would be unduly long, unwieldy and even unnecessary. I presume what

the reviewers would like would be a concise summary that reveals whether the work had produced meaningful results which are leading us to an understanding of the origin of the genetic code and the process of protein synthesis. To that end I will use a tabular form of presenting the conclusions and showing the flow of concepts. The listings will not necessarily be in chronological order. The publications relevant to that particular concept will be listed. Those particular areas which will be the subject of the proposed research will be separately reviewed in some detail just before the proposed work is described.

Table 1. Summary of Previous Research

Type of experiment

Thermal proteinoids interacting with homopolyribonucleotides

References

Fox, S.W., Lacey, J.C., Jr. and Nakashima, T., Interactions of thermal proteinoids with polynucleotides, in Nucleic-protein interactions - Nucleic acid synthesis in viral infections, D.W. Ribbons, J.F. Wosener and J. Schultz, eds. North-Holland Pub. Co., Amsterdam, London (1972) pp 113-127.

Lacey, J.C., Jr., Stephens, D.P., and Fox, S.W., Selective formation of microparticles by homopolyribonucleotides and proteinoids rich in amino acids. Biosystems 11 9-17 (1979).

Lacey, J.C., Jr., Yuki, A., and Fox, S.W., Co-precipitation of thermal lysine rich proteinoids with polyribonucleotides. Biosystems 11 1-7 (1979).

Comments

By and large thermal proteinoids rich in one amino acid and containing 15 mole % lys will coprecipitate selectively with their anticodonic homopolyribonucleotide. Our first indications of anticodonic preferences by amino acids were with thermal proteinoids. Lys-rich preferred poly U, Gly-rich preferred poly C, Pro-rich preferred poly G. The lys preference for poly U is the only data we have for a hydrophilic amino acid preferring its anticodonic nucleotide. Most of the data we have are for hydrophobic amino acids.

Type of experiment

Correlations of properties

References

Lacey, J.C., Jr. and Weber, A.L., The origin of the genetic code: An amino acid anticodon relationship, in Protein Structure and Evolution, L. Fox, Z. Deyl and A. Blasej, eds., Marcel Dekker (1976) pp. 213-22.

Weber, A.L. and Lacey, J.C., Jr., Genetic code correlations: Amino acids and their anticodon nucleotides. J. Mol. Evol. 11 199-210 (1978).

Comments

The preferential binding of proteinoids to their anticodon led us to look for other data. It was found that hydrophobicities or hydrophilicities of amino acids correlate with those of their anticodons. This is good evidence that the code is not random. We began to think of the genetic anticode and the logic of the anticode became more apparent. Some anticodonic assignments did not fit the correlations and we postulated that these assignments were made late in evolution after the process of protein synthesis had become established. This led to a study of tRNA evolution later.

Type of experiment

Selective interaction of amino acids and nucleotides using NMR

References

Khaled, M.A., Mullins, D.W., Jr., Swindle, M. and Lacey, J.C., Jr., Complexes of polyadenylic and the methyl esters of amino acids, Orig. Life 13 87-96 (1983).

Khaled, M.A., Mullins, D.W., Jr., and Lacey, J.C., Jr., Binding constants of phenylalanine for the four mononucleotides, J. Mol. Evol. 20 66-70 (1984).

Khaled, M.A., Watkins, C.L. and Lacey, J.C., Jr., ^2H NMR demonstration of amino acid - nucleotide interactions, Biochem. Biophys. Res. Comm. 106 1426-1434 (1982).

Comments

The correlations of properties must be based on some more fundamental reality. One possibility is that they are due to selective interactions between amino acids and anticodons. In studying that possibility we have used NMR. We found that Phe preferentially associates with its anticodonic nucleotide AMP > GMP > CMP > CMP. Also there is an heirarchy of amino acids (having A as middle anticodonic letter) in their association with poly A; Phe > Leu = Ile > Val > Gly. Gly was included for comparison. In these studies a principle was displayed. If one takes a hydrophobic moiety and studies the binding of a series of second type of compound to the first, the binding constants vary directly as the hydrophobicity of the second type of compound.

Type of experiment

Structural studies of aminoacyl nucleotides and nucleotides

References

- Sternylanz, H., Subramanian, E., Lacey, J.C., Jr. and Bugg, C.E., Interactions of hydrated metal ions with nucleotides; the crystal structure of barium adenosine 5'-monophosphate heptahydrate, *Biochemistry* 15 4797-4802 (1976).
- Lacey, J.C., Jr., Hall, L.M., Mullins, D.W., Jr., and Watkins, C.L., Chirally selective intramolecular interaction observed in an aminoacyl adenylate anhydride, *Orig. Life* 16 151-156 (1985).
- Lacey, J.C., Jr., Mullins, D.W., Jr., and Watkins, C.L., Aliphatic amino acid side chains associate with the "face" of the adenine ring, *J. Biomol. Struct. and Dynam.* 3 783-793 (1986).
- Lacey, J.C., Jr., Thomas, R.D., Hawkins, A. and Watkins, C.L., Differential distribution of D- and L-amino acids on the 2' and 3' positions at the 3' terminus of tRNA, submitted to *Proc. Natl. Acad. Sci.*

Comments

Affinities between amino acids and nucleotide bases should cause certain preferred conformations. NMR studies have shown that with the hydrophobic nucleotide 5'-AMP there are definite intramolecular interactions with hydrophobic amino acids or carboxylic acids either esterified to the ribose group or attached to the phosphate group. The interactions are between the hydrophobic amino- or carboxylic acid and the adenine ring. These interactions tend to give definite conformations to these important compounds and indicate the importance of hydrophobicity as an organizing principle in the origin of the code and the process of protein synthesis. With the adenylate anhydrides there is a preferential interaction of the L-phenylalanine (more than the D isomer) with the adenine ring. Also in esters of the ribose of 5'-AMP, the L-amino acids distribute to the 3' position 67%, 2' 33%. The D-amino acids and carboxylic acids distribute to the 3' position inversely as a function of the hydrophobicity of the acid. This data possibly explains the origin of the exclusive use of the 3' position at the end of tRNA for peptide bond formation and offers at least a partial explanation for the origin of the preferential use of L-amino acids in protein synthesis.

The X-ray structure of Ba-AMP showed the metal was bound to the nucleotide through seven of the eight waters of hydration of the barium. The nucleotide was in the anticonformation with an unusual C_4 (exo) ribose ring.

Type of experiment

Chemical reactions relevant to protein synthesis

References

- Lacey, J.C., Jr. and White, W.E., Jr., Aminoacyl transfer: Chemical conversion of an aminoacyl adenylate to an imidazolide, *Biochem. Biophys. Res. Comm.* 47 565-573 (1972).
- White, W.E., Jr., Lacey, J.C., Jr. and Weber, A.L., Aminoacyl transfer from an adenylate anhydride to the 2'OH groups of polyribonucleotides, *Biochem. Biophys. Res. Comm.* 51 283-292 (1973).
- Weber, A.L. and Lacey, J.C., Jr., Aminoacyl transfer: Peptide synthesis and other properties of an aminoacyl imidazolide, *Biochem. Biophys. Acta* 349 226-234 (1974).
- Weber, A.L. and Lacey, J.C., Jr., Aminoacyl transfer from an adenylate anhydride to polyribonucleotides, *J. Mol. Evol.* 6 309-320 (1975).
- Mullins, D.W., Jr. and Lacey, J.C., Jr., Studies on the chemical basis for the origin of protein synthesis: Initiation and direction of growth, *J. Mol. Evol.* 15 339-345 (1980).
- Mullins, D.W., Jr. and Lacey, J.C., Jr., Genetic code correlations: Differential rates of non-enzymatic activation of hydrophobic amino acids by ATP, *Biochem. Biophys. Res. Comm.* 96 491-497 (1980).
- Mullins, D.W., Jr. and Lacey, J.C., Jr., Phosphate production and analysis in non-enzymatic activation of amino acids by ATP, *Zeitschrift f. Naturforsch.* 36 ab 732-734 (1981).
- Mullins, D.W., Jr. and Lacey, J.C., Jr., Factors influencing the rate of non-enzymatic activation of carboxylic and amino acids by ATP, *Orig. Life*, D. Reidel Pub. Co. (1981) 209-215.
- Mullins, D.W., Jr., Senaratne, N. and Lacey, J.C., Jr., Aminoacyl-nucleotide reactions: Studies related to the origin of the genetic code and protein synthesis, *Orig. Life* 14 597-604 (1984).
- Mullins, D.W., Jr. and Lacey, J.C., Jr., Highly efficient peptide formation from an N-acetyl aminoacyl-AMP anhydride and a free amino acid, *J. Mol. Evol.* 19 176-178 (1983).
- Lacey, J.C., Jr., Senaratne, N. and Mullins, D.W., Jr., Hydrolytic properties of phenylalanyl and N-acetyl phenylalanyl adenylate anhydrides, *Orig. Life* 15 45-54 (1984).
- Lacey, J.C., Jr., Hall, L.M. and Mullins, D.W., Jr., Rationalization of some genetic anticodonic assignments, *Orig. Life* 16 69-79 (1985).
- Lacey, J.C., Jr. and Mullins, D.W., Jr., Genetic coding catalysis, *J. Theo. Biol.* 115 595-601 (1985).
- Mullins, D.W., Jr. and Lacey, J.C., Jr., Comparative rates of esterification of 5'-AMP with hydrophobic amino acids: Relevance to genetic code assignments, *J. Mol. Evol.* 23 76-79 (1986).

Comments

The studies in these papers have shown that virtually all of the steps in protein synthesis are spontaneous and some are selective. The studies include non-enzymatic activation of amino acids with ATP, transfer of amino acid from adenylate anhydride to imidazole and then to 2'OH groups of polyribonucleotides or 2'-3' groups of 5'-AMP, formation of peptides from imidazolides and from adenylate anhydrides, selective (Phe > Leu > Val > Ile) esterification rates of 5'-AMP and poly'A using aminoacyl imidazolides.

The hydrolytic properties of amino acyl-, N-blocked aminoacyl- and carboxylic acid adenylate anhydrides and esters were determined. Generally N-blocked and carboxylic acid derivatives are more stable than free aminoacyl derivatives. Comparative rates of esterification of 5'-AMP and 5'-UMP with Ac-Phe-imidazolidine showed the anticodonic nucleotide (5'-AMP) reacted faster. This latter finding led to an hypothesis that the amino acid side chains and base rings are like "piggy back" catalysts giving built in code-related selectivity in amino acid-nucleotide reactions.

Reviews, Summaries and Speculative Papers

References

- Lacey, J.C., Jr. and Fox, S.W., Concepts related to the origin of the genetic apparatus, in *Theory and Experiment in Exobiology*, A. Schwartz, ed. Wolters-Noordhoff Pub., The Netherlands 2 35-63 (1972).
- Lacey, J.C., Jr. and Mullins, D.W., Jr., Proteins and nucleic acids in prebiotic evolution, in *Molecular Evolution! Prebiological and Biological*, D.L. Rohlfsing and A.S. Oparin, eds., Plenum Press, N.Y. (1972) 171-188.
- Lacey, J.C., Jr. and Mullins, D.W., Jr., Isolated microsystems in evolution, in *Origin of Life and Evolutionary Biochemistry*, K. Dose, S.W. Fox, G.E. Deborin and T.E. Pavloskaya, eds., Plenum Press, N.Y. (1974) 311-319.

Page intentionally left blank

Page intentionally left blank

One might summarize the results presented in the table by saying that there is little doubt that there was a definite molecular basis for the origin of the bulk of the genetic code assignments. This is principally revealed by the correlations of properties of amino acids and their anticodons. However, the idea is further supported by biases in binding constants between amino acids and nucleotides and these in turn give rise to biases in reaction rates between amino acids and nucleotides. The data not only offer beginning explanations for distinguishing between different amino acids, but also between the D- and L-forms of the amino acids. Furthermore, intramolecular interactions between amino acids and nucleotide bases in aminoacyl nucleotide anhydrides and esters give rise to specific conformations of these molecules, which conformations in turn can play a role in the self organization of molecular systems. The prognosis remains good, I believe, that study of the character of the component molecules will allow us to understand the basis of the origin of those biochemical entities, the genetic code and protein synthesis.

Proposed Program

Because the foregoing summary of all of the previous work was so abbreviated, additional background information will be furnished along with each of the restricted categories of proposed research so the reviewers can be aware of the extent and details of previous work and what remains to be done in each case.

2'-3' distribution of amino- and carboxylic acids when esterified to nucleotides

Every amino acid coming into protein synthesis is esterified to the ribose group of the AMP at the 3' end of its cognate tRNA. Although each aminoacyl

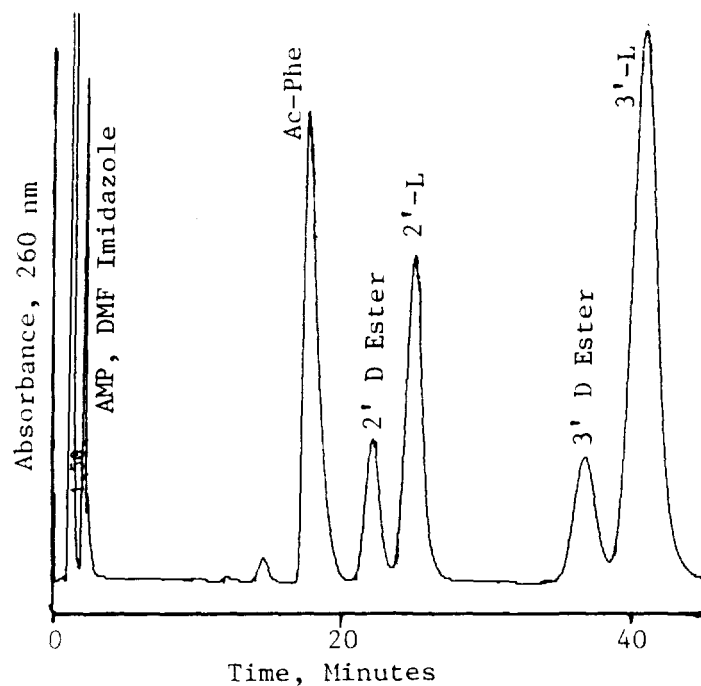


Fig. 1 HPLC tracing for separation of crude Ac-Phe-AMP ester on phenyl Bondapak column (5 μ particles; 3 mm column 30 cm long) using 13% methyl alcohol, 0.05 M phosphate, pH 2.0, 1.0 ml/min; chart 3 mm/min. Absorbance at 260 nm with a range of 0.05. Peaks were identified by PMR. Note that in this preparation beginning with Ac-L-Phe, some Ac-D-Phe AMP ester is present due to partial racemization during the reaction. Also note that the 3' Ac-L-Phe-AMP peak is about twice the size of the corresponding 2' Ac-L-Phe-AMP peak, but that the two D isomers are nearly the same size.

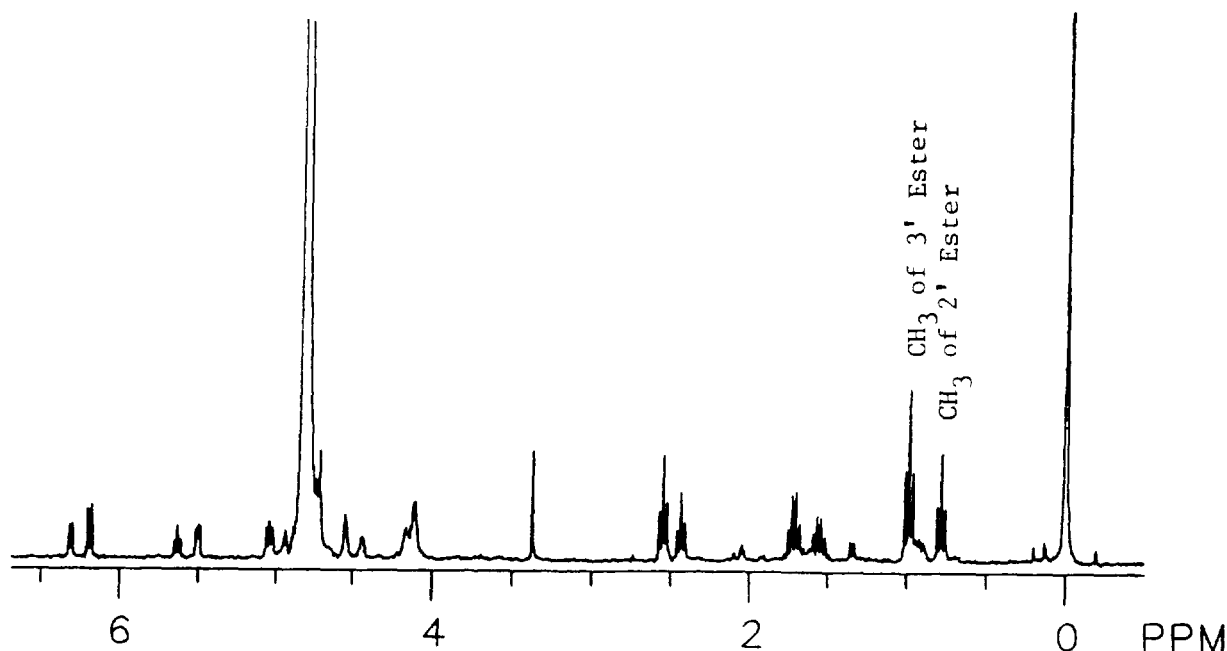


Fig. 2 PMR spectrum of butyric acid-AMP ester at 0.004 M in 0.05 M phosphate and pD 6.5 using our high resolution Nicolet 300 NMC Spectrometer. The spectrum shown is after about 200 accumulations. The reference is external deuterated TSP (Sodium-3-trimethylsilylpropionate-2,2,3,3- d_4). The proton peaks for the H_2' and the H_3' of the corresponding butyric AMP esters are shown downfield at about 5.5 ppm and are integrated to give the 1'3' ester. The separate H_1' doublets at about 6.3 ppm can also be integrated to give the 1'3' ester. Upfield signals representing the butyric chain also show two sets of signals, one for the 2' ester and one for the 3' ester. In every case the 2' ester peaks are further upfield, suggesting that when in the 2' position, the hydrophobic butyric chain can associate more with the adenine ring than when in the 3' position.

"Page missing from available version"

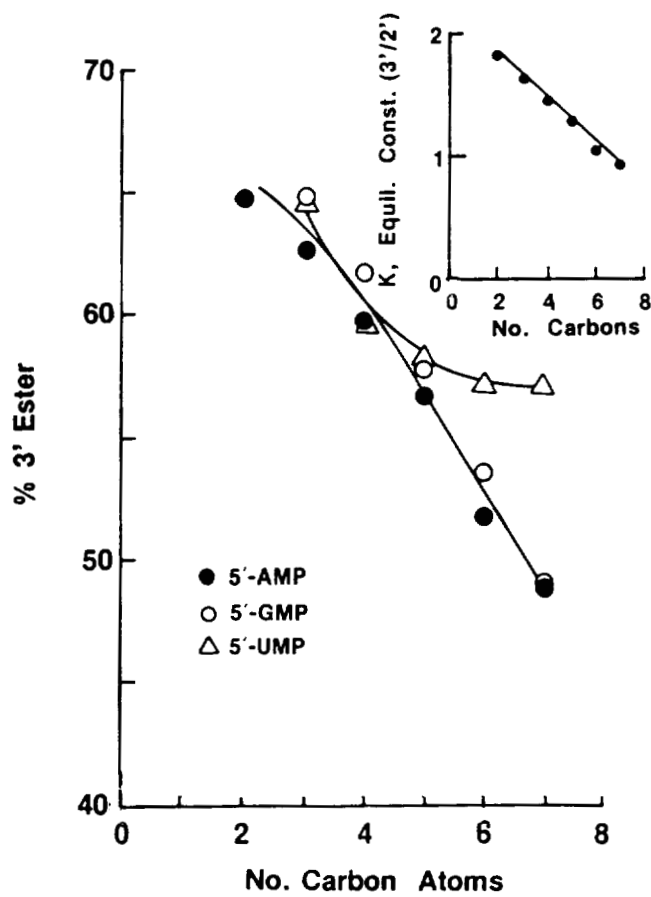


Fig. 4 Percent 3' ester as a function of the length of the carboxylic acid chain length (linear chains only) for esters of 5'-AMP, 5'-GMP and 5'-UMP. Inset is the data for 5'-AMP only expressed as the equilibrium constant ($\% 3' / \% 2'$).

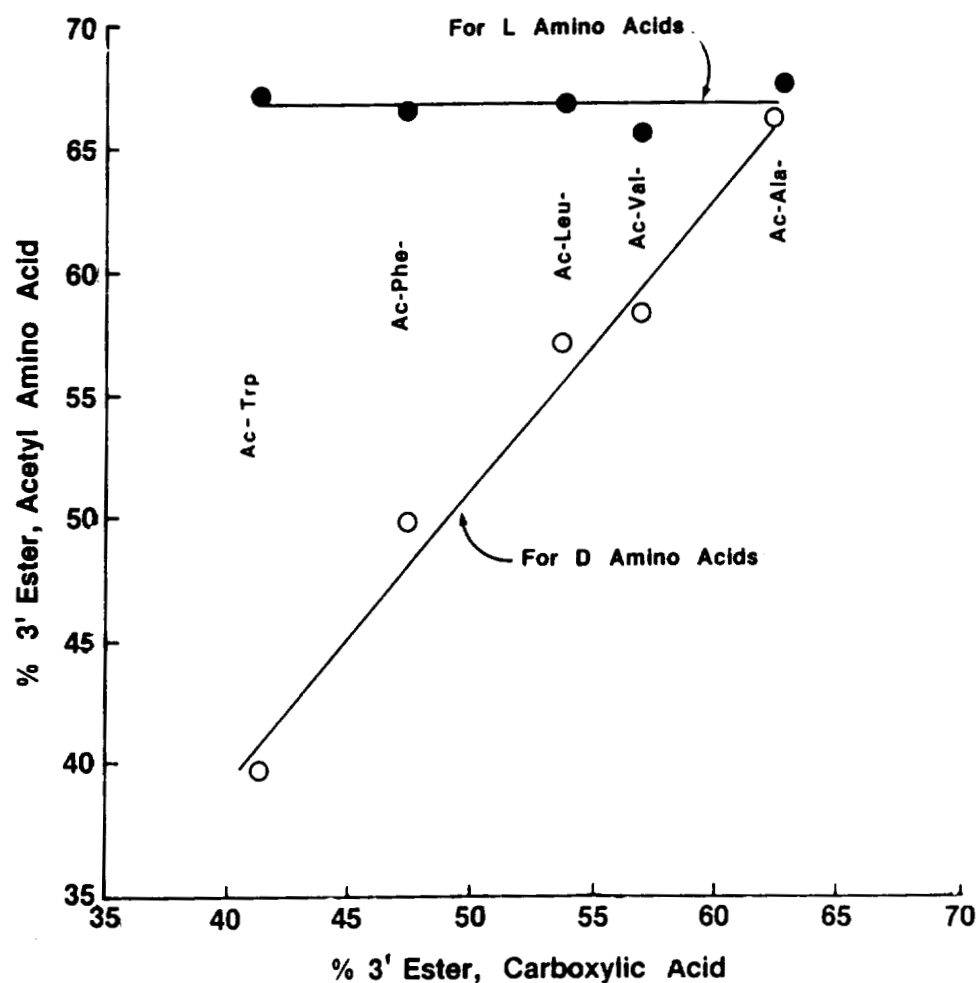


Fig. 5 Percent 3' ester of the Ac-D- and L-amino acids plotted versus the data for their carboxylic acid analogs. These data are for 5'-AMP esters only. The analogs are propionic (for ala); isovaleric (for val); 4-methyl valeric (for leu) and phenyl-3-propionic (for phe). Because the slope of the line for the D-amino acids is about 45°, it shows that the D-amino acids are behaving identically as their carboxylic analogs. The L-amino acids are distributing ~ 67% to the 3' position regardless of the side chain.

The results with 5'-GMP so far are almost the same as with 5'-AMP (Table 2). However, the results with 5'-UMP show that hydrophobic carboxylic acids are not discriminated against in the 3' position as much as with 5'-AMP and 5'-GMP (Fig. 4).

Because the pyrimidines are much less hydrophobic than the purines (Weber and Lacey, 1978), one would expect the pyrimidines to be less effective, assuming hydrophobic interactions are playing a role. If it turns out that the two pyrimidines are less discriminating against the D-amino acids, then it may explain another aspect of contemporary tRNA composition. The fourth nucleotide from the 3' end of tRNA is variable (Crothers et al., 1972) but survey of the 1985 compilation (Sprinzl et al., 1985) shows that 71% of the sequences have an A in that position, 20% have a G with the pyrimidines totalling only 9% together. This observation suggests that primitive tRNAs may have existed which terminated principally in either A or G instead of -CCA, the purines being there to help discriminate against hydrophobic D-amino acids.

Our proposed research on this category is to add more hydrophilic amino acids to the 5'-AMP data, to acquire the same complete set of data with 5'-GMP, 5'-CMP and 5'-UMP using synthesis, HPLC and NMR procedures already described in Table 2. In essence we propose to acquire the data to fill in the blank spaces in Table 2.

When the results are complete we should be able to draw more firm conclusions regarding the molecular basis for the origin of the preferential use of the L-amino acids in the 3' position for protein synthesis. Secondly, the results from all four nucleotides may allow us to say why there is a preference

Table 2. Percent 3' Isomer for Various Acetyl Amino Acids and Carboxylic Acid Esterified to 5'-AMP, 5'-GMP or 5'-UMP

Amino or Carboxylic Acid	5'-AMP	5'-GMP	5'-UMP	5'-CMP
Ac-Gly	68.3 (66.7)	71.3 -		
Ac-L-Arg	- -	- -	- -	- -
Ac-D-Arg	- -	- -	- -	- -
Ac-L-Lys	- -	- -	- -	- -
Ac-D-Lys	- -	- -	- -	- -
Ac-L-Asn	- (68.8)	- -	- -	- -
Ac-D-Asn	- (62.0)	- -	- -	- -
Ac-L-Ala	68.7 (65.9)	68.6 -	- -	- -
Ac-D-Ala	69.9 (67.9)	- -	- -	- -
Ac-L-Pro	66.7 -	- -	- -	- -
Ac-D-Pro	- -	- -	- -	- -
Ac-L-Val	66.6 -	66.0 -	- -	- -
Ac-D-Val	58.5 (59.6)	59.9 -	- -	- -
Ac-L-Leu	65.4 (67.1)	69.7 (68.1)	- -	- -
Ac-D-Leu	56.0 (56.0)	55.0 (59.2)	- -	- -
Ac-L-Phe	66.1 (66.0)	67.1 (67.1)	73.0 -	65.0 (67.3)
Ac-D-Phe	47.5 (45.8)	46.1 -	- -	- (49.9)
Ac-L-Trp	67.6 (65.1)	- -	- -	- -
Ac-D-Trp	41.0 (37.6)	- -	- -	- -
Acetic	64.9 (65.0)	- -	- -	- -
Propionic	62.6 (63.5)	64.8 -	64.8 -	- -
Butyric	59.8 (60.3)	61.8 -	59.7 -	- -
Valeric	56.6 (56.8)	57.8 -	58.2 -	- -
Isovaleric	57.0 (56.5)	- -	51.0 -	- -
3-Me Valeric	53.7 -	- -	50.3 -	- -
4-Me Valeric	54.0 (54.2)	- -	50.9 -	- -
Caproic	51.7 (51.7)	53.2 -	57.6 -	- -
Heptanoic	48.9 (47.6)	48.7 -	56.1 -	- -
Phenylpropionic	47.5 (49.5)	- -	58.5 -	- -

Each value is the average of at least three separate determinations by integration of HPLC peaks. Numbers in parentheses are confirmatory experiments using proton NMR.

The HPLC procedure was as follows: A sample of crude ester at pH 2.0 was injected into the apparatus (Waters 3mm, 30cm, C₁₈ Bondapak, 5 particles) and with a pumping rate of 0.8-1.0 ml/min. The products were eluted with 0.05 M phosphate, pH 2, containing different amounts of HPLC grade methanol for example (8% Ac-Ala; 15% Ac-Val; 20% Ac-Leu and Ac-Phe esters). A portion of the larger product peak (3' ester) was collected and lyophilized to remove methanol, reconstituted and incubated at 25 °C and pH 7.0 for 20 min to allow 2'-3' equilibration. The pH was again dropped to 2.0 with 4 N HCl and the sample was again run through the HPLC integrating the 2' and 3' peaks. From the relative integrations, the % of each isomer and the equilibrium constant (%3'/%2') were calculated. Three samples were run for each amino acid and the Student T test showed (except on Ac-Ala) that the D and L samples having $p = < 0.05$, i.e. $> 95\%$ probability that the D and L populations are different.

For the proton NMR procedure, crude samples were run through a preparative C₁₈ reverse phase column (2 columns in series, each is 7.8 mm x 61 cm with 50 packing) using 0.001 phosphate buffer at pH 2.0 with varying amounts of methanol for the different esters (30-35% for Ac-Phe and Ac-Leu, 20% for Ac-Val and 15% for Ac-Ala esters) with a pumping rate of 2.0 ml/min. After free AMP, amino acid imidazole and reaction solvent (DMF) were eluted, the combined 2' and 3' esters were collected. The samples were adjusted to pH 6.5 and lyophilized to dryness and reconstituted to a concentration of 3-4 ml in D₂O, lyophilized again and dissolved in D₂O to give a concentration of 0.004 M at pD 6.5 (pD = pH + 0.4) and maintained at -70 °C until the PMR assay. The NMR was done on a Nicolet NMC 300 in the pulse mode accumulating sufficient scans to give clear signals of the H₁', H₂' and H₃' protons generally in the range of 5.4-6.2 ppm. The H₁', H₂' and H₃' protons were all integrated allowing two independent estimates of the % 3' ester (there is a separate H₁' signal for the 2' and the 3' ester). Only one PMR sample for each was run because we mainly wanted to confirm that the HPLC peaks had been given the proper identity.

for purines (and principally A) in the fourth nucleotide from the 3' end of tRNAs.

Structural Studies of Aminoacyl Esters

The work proposed in the previous section would tell us differences in 2'-3' distribution of the D- and L-amino acids and carboxylic acids. Those observations are of interest of their own right. However, it would be even more desirable to understand the basis for variations in the 2'-3' distributions. Why do all of the L-amino acids distribute 67% to the 3' position? Why are the more hydrophobic D-amino and the carboxylic acids drawn toward the 2' position? In the process of trying to answer these questions, we might also discover the molecular basis for other aspects of protein synthesis. For example, in the previous section it was suggested that hydrophobic D-amino acids might be drawn toward the 2' position because in that position their side chains associate more closely with the hydrophobic adenine ring. We have previously noted intramolecular association of both aliphatic and aromatic side chains with the adenine ring in the aminoacyl adenylate anhydrides (Lacey, et al., 1986) and Usher and Needels, 1984c) had even before reported the association of alanine methyl with the "face" of the adenine ring. If these associations are indeed taking place, then they can also be responsible for longer range intermolecular associations in which linear arrays of aminoacyl nucleotides are formed. Such arrays can lead to complexes which are more likely to form peptides. Using NMR one can study intramolecular interactions at low concentrations (e.g. 4 mmolar) and intermolecular associations at high concentrations (e.g. 40 mmolar).

What we would like to do is to use one or two of these compounds and go after the structure using NMR and X-ray diffraction. Because the intramolecular

interactions are stronger between adenine and Phe or Trp, we propose to study them first.

Before tackling the amino acids, we propose to try the carboxylic acid analogs, phenyl-3-propionic acid and indole-3-propionic acid, because we don't have to worry with the D and L forms. Furthermore, even though we would like to eventually know the structures of the esters, we probably will start with the anhydrides (carboxylic acid attached to the phosphate) to avoid the 2'-3' problem, therefore, there is only one compound to deal with. The program would be as follows:

1. Synthesize the AMP anhydrides of indole-3-propionic acid and phenyl-3-propionic (analogs of Trp and Phe) using methods we have described previously (Lacey, et al. 1984).

2. Purify them using a combination of acetone precipitation and reverse phase HPLC. These procedures have already been worked out using a Waters phenylbondapak column and our Laboratory Data Control HPLC unit.

3. Using a combination of simple PMR and NOE experiments, determine as nearly as possible which protons in the carboxylic acid rings are associating with the adenine ring. The program for the NOE has already been set up and satisfactorily employed using 3-methyl estrone as a test compound by irradiating the 3-methyl protons and looking at the effects on the protons at positions 2 and 4. NOE enhancement of the two signals was observed. With the adenylate derivatives, it is not possible to predict which protons should be irradiated. More than likely, the H-8 proton of the adenine will be most involved, but experiment will have to dictate which phenyl- and indole- protons are most

appropriate. Because we are interested in intramolecular interactions we must run at low concentrations, consequently 2 D NOE or NOSY experiments may not be possible due to the expense because so many accumulations are required.

We will also attempt to crystallize the barium salts of these two carboxylic acid anhydrides. We have previously studied the hydrolytic stability of a series of carboxylic acid - AMP anhydrides and find they are quite stable even at pH 7 ($k = 1 \times 10^{-4}/\text{min}$). The hydrolysis constant is an order of magnitude lower than that for Ac-Phe-AMP anhydride. So these compounds are stable enough to allow time for crystallization, especially if the temperature is kept low. The major objective here would be to understand the intramolecular interactions of the hydrophobic side chains of the carboxylic acids with the hydrophobic adenine ring. Our only role in this would be synthesis, purification and crystallization. The X-ray diffraction studies would be carried out by Dr. Charles Bugg's group at no charge to our grant. We have collaborated with Dr. Bugg before in this same way to determine the structure of the barium salt of 5'-AMP (Sternglanz et al., 1976).

Dr. Bugg is head of our X-ray diffraction group and is associate research director of our Comprehensive Cancer Center. A letter from him agreeing to collaborate is enclosed. We can send a copy of his lengthy curriculum vitae if the reviewers desire.

The research plan would then be to move to the phenyl-3-propionic and indole-3-propionic esters of 5'-AMP. We have also already synthesized (Gottikh et al., 1970) and purified (using HPLC) these compounds recently. Again, the plan for studying their conformation would involve NMR and crystallography as proposed for the anhydrides above.

We would then move to esters of Ac-D- and L-Phe AMP and Ac-D- and L-Trp AMP. We have experience with preparation of these compounds (using the aminoacyl imidazolid method of Gottikh et al., 1970) and their purification using reverse phase (phenyl bondapak) chromatography. That previous work was for the 2'-3' distribution study. For NMR and crystallography we need larger quantities but we do not anticipate any real difficulties. The NMR and crystallography studies would be carried out as described for the compounds above.

The main objective with these D- and L-aminoacyl esters is to understand:

1. Why do all of the L-amino acids distribute ~ 67% to the 3' position of AMP?
- and 2. Why do the more hydrophobic D-amino acids tend to be drawn to the 2' position?

As mentioned previously, the results we have thus far suggest that hydrophobic D-amino acids and carboxylic acid analogs behave similarly in their 2'-3' distribution on 5'-AMP. By looking at the more simplified PMR spectra of a model compound like butyric acid ester of 5'-AMP (Fig. 2) we can see that when the butyric acid is in the 2' position the signals for the butyric acid side chain protons are shifted upfield relative to those of the 3' ester. Consequently, as a first explanation we can say that hydrophobic interactions of the carboxylic acid side chains and hydrophobic amino acid side chains with the adenine ring are greater when these acids are in the 2' position. As a result of the proposed NMR and X-ray crystallography, we would hope to come to a more precise explanation as to just where these intramolecular interactions are taking place. These studies are of particular interest regarding the origin and evolution of processes involving the 3' aminoacylated tRNA however, they are also of importance regarding our long-standing interest in the way that amino acids and nucleotides interact with each other.

Oligopeptide-oligonucleotide interactions

This particular subject was part of an earlier proposal, however, work was delayed on it because of the unexpected and interesting results with the 2'-3' distribution study which we have concentrated on for the last year and a half. Furthermore, we had some difficulty in purifying the oligonucleotides which were furnished us. Basically, the proposed program was to study the interaction of a tetrapeptide (phe-gly-lys-pro) with its codonic and anticodonic strands. These particular amino acids were chosen because they are the so-called homocodonic amino acids, having homogeneous codons, UUU, GGG, AAA and CCC respectively. The plan was to purchase the peptide and oligonucleotides, purify them and then study their interaction using NMR. We did purchase 1.0 gram of the peptide and purified it to > 97 % purity using HPLC. We have run PMR on it and assigned all of the peaks (Fig. 6). However, we had trouble with the oligonucleotides. Even after running through ion exchange FPLC, even though there is only one band in reverse phase chromatography, we consistently get two bands when running a sequencing gel (P^{32} label). Difficulty with purification is apparently a common problem with synthetic oligonucleotides. Dr. Brian Johnson in an adjoining lab is purifying oligonucleotides using preparative gel electrophoresis. He has offered to help us with the procedure.

We do have on hand, 50 O.D. units of the dimeric repeats of the deoxyoligonucleotides UUGGAACC and GGUUCCAA and as soon as these are successfully purified we will begin the NMR study. We must first assign all of the peaks in the oligonucleotides. Then we will study the binding constants by titration of the oligonucleotides with the peptide. We have successfully employed this technique previously (Khaled, et al., 1983; Khaled, et al., 1984) by following changes in

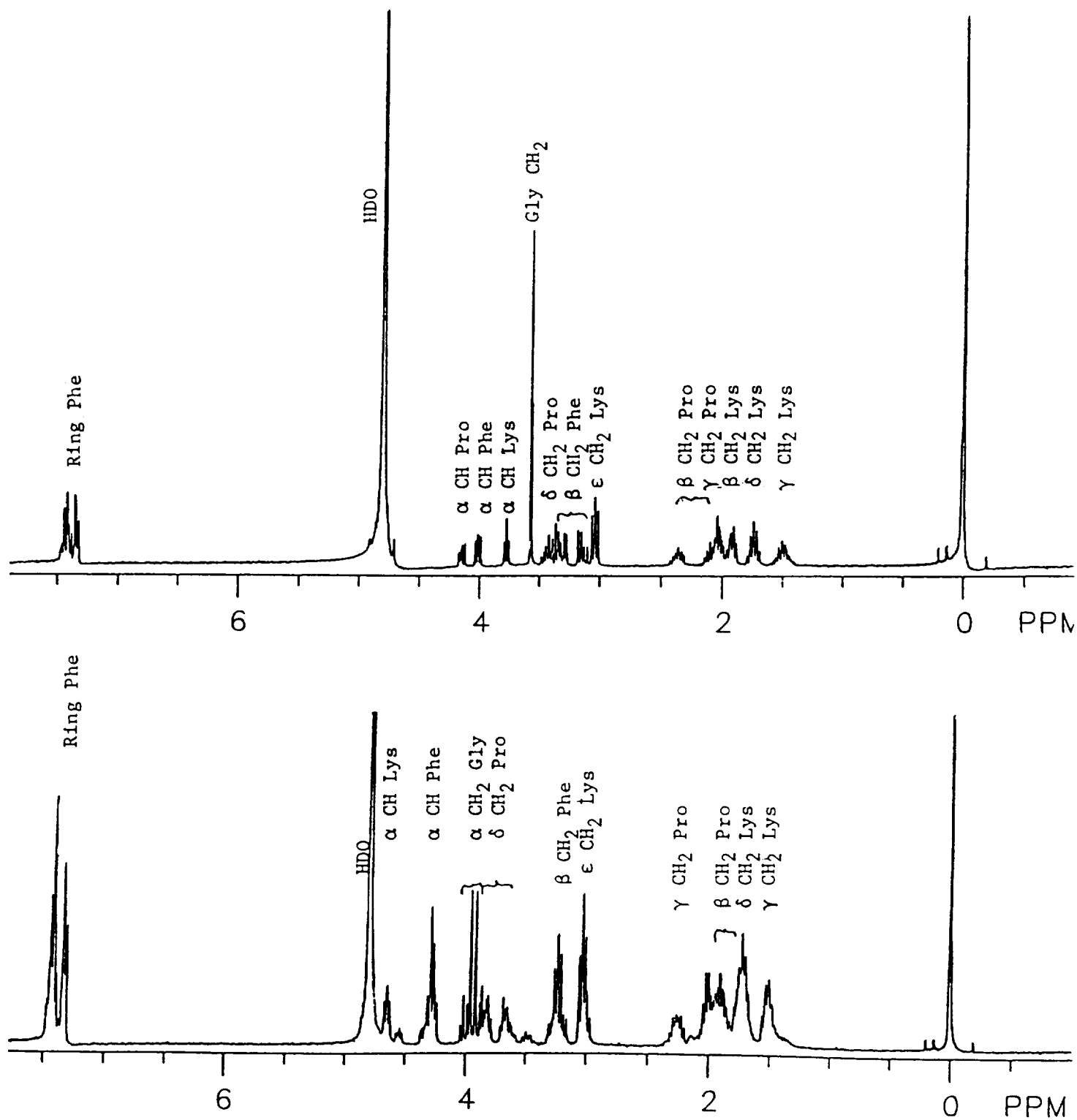


Fig. 6 Proton NMR spectra of A. an equimolar mixture of Phe, Gly, Lys and Pro and B. a 0.004 M solution of the tetrapeptide Phe-Gly-Lys-Pro at pH 7.0 in 0.05 M phosphate using TSP as an external reference.

chemical shifts of selected protons. Such titrations normally result in a saturation curve when plotting $\Delta\delta$ vs. concentration. A linear plot can then be obtained by plotting $\Delta\delta/\text{concentration}$ vs. $\Delta\delta$. The slope of the resulting plot gives the binding constant in reciprocal moles. We believe we can follow changes in the position of the adenine protons due to interactions with the phe group in the peptide. We have already documented ring current effects of phe on adenine protons (Khaled et al., 1983; Khaled et al., 1984) including the preferential association with adenine (Khaled et al, 1983).

We have recently acquired a used ORD-CD apparatus (JASCO UV 5) from Dr. Sidney Fox's laboratory and will also attempt to study the peptide binding to the oligonucleotides using this equipment. Formation of complexes can often cause dramatic changes in the conformations of peptides and oligonucleotides. These conformational changes will result in changes in the ORD and/or CD spectra. Again a peptide titration of the oligonucleotide would be used, following the resulting changes, plotting the $\Delta/\text{concentration}$ peptide vs. Δ gives a straight line the slope of which is the binding constant in reciprocal moles.

The ORD/CD work will be contingent on our being able to get the apparatus working after the trip in a U-Haul truck from Miami. There is every reason to believe that we will be successful in that.

After studying the dimeric repeats, we will proceed to the trimeric repeats, 5'-UUUGGGAAACCC-3' and 5'-GGGUUCCCAAA-3' and proceed through the same steps outlined above for the dimeric repeats. Again, the major question is, "Which strand is preferred by the tetrapeptide, the codonic or anticodonic?"

These studies can not only help answer the fundamental question of codonic or anticodonic preference, but also whether the 3/1 coding ratio has resulted from a preferred 3/1 ratio regarding the α -helix of peptides and the helical form of the nucleic acids.

Future studies might involve trying to cocrystallize the peptide and oligonucleotides. Such a study would involve much larger quantities of the oligonucleotides, i.e. many milligrams. In trying to elucidate the basic structure of peptides complexes with oligonucleotides however, X-ray diffraction studies of crystals would be the preferred approach.

The NMR work will be carried out by postdoctoral fellow Ralph Thomas (curriculum vitae enclosed) in collaboration with Dr. Charles L. Watkins of our Chemistry Department (letter of intent and curriculum vitae enclosed) who has collaborated with us a number of years now and co-authored several papers with us. The NMR instrument is a Nicolet 300 MHz high resolution spectrometer that has been used in all of our past NMR studies.

Peptide formation

Weber and Orgel (1980,1981) have shown that the yield of glycine peptides can be increased, when starting with glycine-GMP (codonic combination) or glycine-AMP ester, if a poly C or poly U template is used. The improvement was small. We would like to repeat their type of experiment only using an anticodonic combination of amino acid and mononucleotide. We propose to study peptide formation from Phe-AMP ester. We already know from our NMR studies that the Phe ring - adenine interaction is taking place in these compounds and as reported above, Ac-L-Phe distributes 67% to the 3' position and Ac-D-Phe only 50% to the

3' position. An interesting question is, "Is there some fundamental property of these esters which gives a higher yield of peptides from the L-amino acid in the 3' position?"

When an active aminoacyl nucleotide, such as Phe-AMP ester, is put into aqueous solution, it is subject to hydrolysis as well as peptide bond formation. We already have shown that at 10^{-4} M, and lower, hydrolysis is the sole detectable reaction. However, if we study peptide bond formation as a function of concentration, we should see a gradual increase in peptide yield as a function of concentration. This is mainly due to the increased probability of two esters colliding as concentration increases.

However, in addition, there should be additional factors at work involving successful positioning in the transition state. We believe that if we study peptide yields as a function of concentration, at some point stacked arrays of Phe-AMP will begin to form as a result of the Phe-adenine interaction. Several things might result from this. The intermolecular interactions might well force a higher % 3' ester with the L-isomer. If this particular conformation is best for peptide yields, then we might see a dramatic increase in peptide yields. Similar experiments with the D-Phe-AMP ester would show which isomer is best suited for peptide bond formation. We can determine whether there are changes in the 2'-3' distribution of the amino acid as a function of concentration by using the NMR technique mentioned in the earlier study on the structure of aminoacyl esters. We can then correlate peptide yields with the % 3' ester and see if there is a relationship there. For studying the complexation, we might use pD 5 (pD = pH + 0.4) to inhibit peptide bond formation. The effect of concentration on the % 3' ester can also, and will, be studied using the HPLC method. Again, we would have to use pH 5 to inhibit peptide formation.

s, Ultimately we will study the effect of complexation of D- and L-Phe-AMP esters with poly U as a template. We have previously determined the conditions necessary for forming the complex. Using 10 mmolar Phe-AMP and 20 mmolar poly U and 40 mmolar Mg^{++} at 5 °C gives a precipitate which contains a 2/1 U/AMP ester ratio. Formation of the complex should allow efficient peptide formation, however, previous efforts at this by us have yielded no peptides. This means, we believe, that we need to study the optimal conditions for peptide formation from the Phe-AMP-ester before we attempt the templating. pH is probably a major variable, but divalent cation concentration could well be another important one. For this study we would, at a fixed concentration, study the effect of pH from pH 5 to 9 on peptide yields with no added cation and then at the optimum pH, study the effect of varying Mg^{++} from zero to 50 mmolar. Peptide yields from all of these studies can be followed using reverse phase HPLC. We have worked out the conditions for detecting sizes up through the tetramer.

Although the experiments we have described so far were with separate D- or L-amino acid isomers, it would be very interesting to study a mixture of D- and L-Phe AMP esters, determining the composition of the peptide products after purifying them with HPLC. We already have a procedure for this using D-amino acid oxidase and analyzing the acid hydrolysate. The procedure requires 1 μ mole of D-amino acid in 1.0 ml H_2O added to 2.0 ml of assay solution containing 20 mg crude D-amino acid oxidase @ pH 8.2 in 0.02 M pyrophosphate. H_2O_2 and catalase may be added to speed up the reaction however, even without their addition, it will go to completion in about 30 minutes. With phenylalanine, one can follow the formation of phenylpyruvate by monitoring the absorbance at 280 nm. The 280 is approximately 0.1 absorbance units per μ mole of D-phe. The basic proce-

dure would be to form the peptides, separate them from AMP and phe using HPLC, lyophilize down, hydrolyze overnight in 6N HCl @ 110 °C in a sealed ampoule, evaporate to remove HCl, reconstitute and analyze for the total phe (ninhydrin assay or HPLC) and for the amount of D isomer using D-amino acid oxidase.

One would hope that some chiral selectivity would be observed in these studies, giving us another piece of information as to why L-amino acids are preferred in protein synthesis.

tRNA evolution

The correlation of hydrophobicity rankings of amino acids and their anticodon dinucleotides (Fig. 7) shows that four assignments, Trp, Tyr, Ile and the XGA anticodons for Ser do not correlate. We had earlier suggested (Lacey and Mullins, 1983) that these assignments were made late in evolution based on mutations of existing tRNAs, or rRNAs, after the process of protein synthesis had been established. If that were true, then these recent tRNAs should show high homologies with the RNA molecules from which they were derived. Furthermore there should be some logic as to which triplets were used in these late anticodon assignments. Regarding the latter point first, if one plots the hydrophobicity rankings of the ordered dinucleotides (Fig. 8) it is obvious that some of the rankings overlap. Particularly noticeable is the overlap of AU (Ile) and GA (Ser) but also CA and UA overlap GU. We are suggesting that because of ambiguity resulting from overlapping properties XAU, XGA, XCA and XUA were not used as anticodons for amino acids early in evolution and were therefore anticodon equivalents of terminator codons. Partially supporting this argument is the fact that XCA and XUA still represent the anticodon equivalents of all contem-

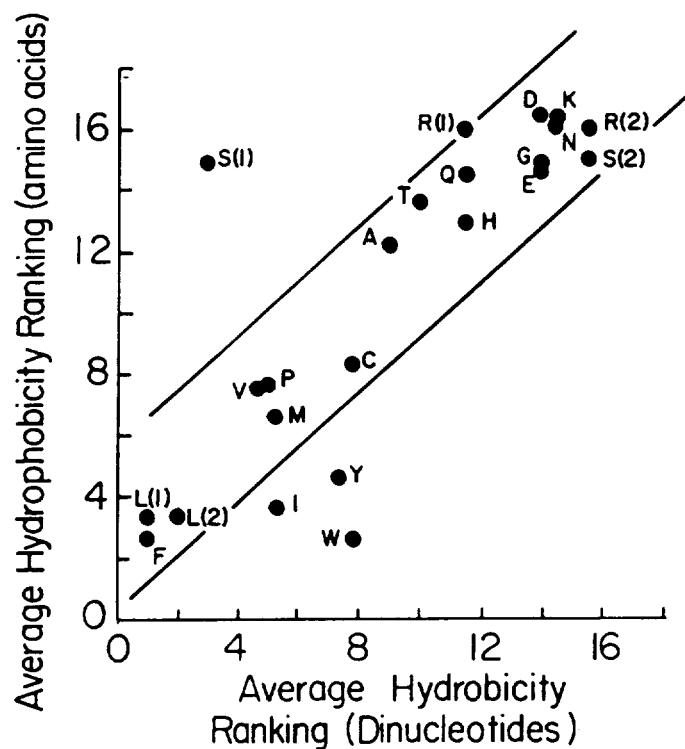


Fig. 7 The average ranked hydrophobicity of the amino acids as a function of the average ranked hydrophobicity of their corresponding anticodon dinucleoside monophosphates representing positions Y and Z of their anticodons XYZ i.e. the anticodon equivalent of the wobble codonic position is not included. The standard single letter abbreviations for the amino acids are used. Amino acids with accompanying numbers are those which possess two sets of anticodon assignments. The correlation coefficient for those points within the outlined band is 0.97 (from Lacey and Mullins, 1983).

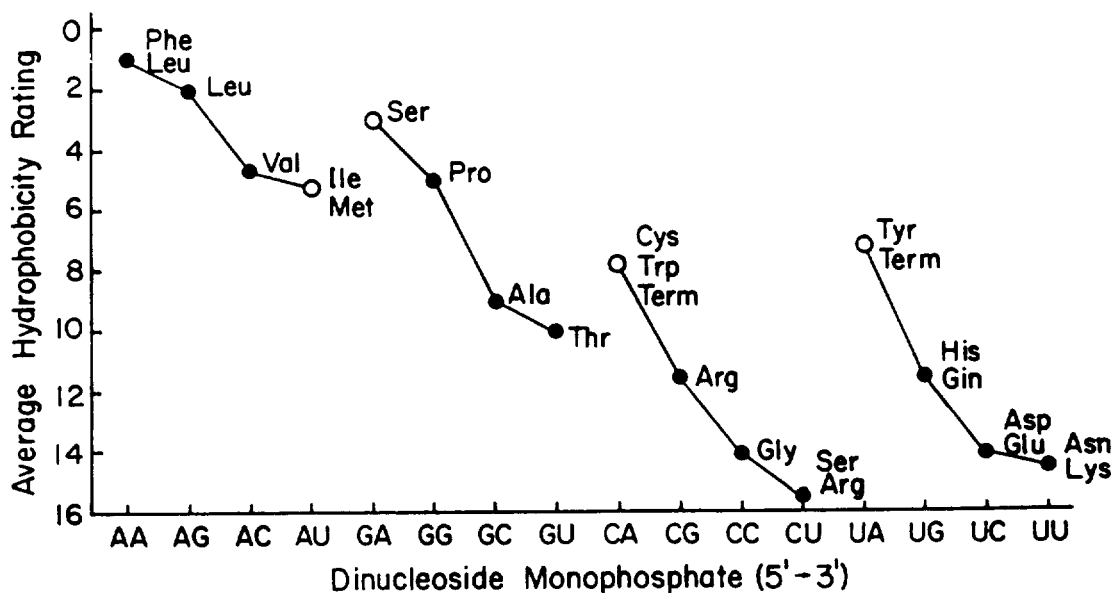


Fig. 8 Average hydrophobicity ratings of the four ordered families of dinucleoside monophosphates. The open circles correspond to those amino acid anticodon doublet assignments which do not fit the correlation of Fig. 7 (ser XGA, trp, tyr and ile).

porary terminators. These particular families then represented groups of unused anticodons which could be assigned at a later date. Because the protein synthesis apparatus was already established, the simplest and most likely source for new tRNAs was through mutation of existing tRNAs, or possibly 5s rRNA. We had earlier shown that one of these, E. coli tRNA^{Tyr} precursor, which is 120 nucleotides long, has quite high homology with E. coli 5s rRNA (Mullins et al., 1973) and we suggested that the two molecules at least had a common ancestor. We subsequently showed that E. coli tRNA^{Ile} 1 and 2 have very high homology with E. coli tRNA^{Val} and E. coli tRNA^{Lys} respectively and suggested that the Ile tRNAs were derived from tRNA^{Val} and tRNA^{Lys} (Staves et al., 1987). We have also shown that E. coli tRNA^{Trp} has high homology with and appears to have been derived from E. coli tRNA^{Gly} 5 (Staves et al., 1987). Consequently, there seems evidence that three of the non-correlating amino acids were given their assignments late in evolution. Further evidence exists regarding Ile. Studies have shown it is very slow to react in the esterification of 5'-AMP (Mullins and Lacey, 1986) and poly A (Lacey, et al., 1985). Furthermore, Ile is one of the two amino acids having two asymmetric centers. Thr is the other one. Therefore, it would seem the evolution of enzymes to synthesize and utilize Ile might have taken more time than most. Regarding Trp and Tyr, they are still two of the least frequently used amino acids (Schwartz et al., 1976).

Mark Staves, who recently received his PhD from the Botany Department, University of Texas, Austin, will be beginning a postdoctoral fellowship with me November 1, 1987. The Department of Biochemistry at UAB will be furnishing the majority of the funds for the first year of this fellowship until October 31, 1988. His curriculum vitae is enclosed.

With Staves' arrival I hope to pursue more vigorously several aspects of tRNA evolution. Firstly, we will study the evolution of E. coli tRNA^{Ser} XGA, the other non-correlating assignment. Comparison will be made with other E. coli tRNAs looking for high homologies. If none are found we will then consider E. coli 5s rRNA as a possible source. Computer programs (Goad and Kanehisa, 1982) that were used for the previous studies will be used for these studies also.

We would then like to consider two more general questions and finalize our work on those. One has to do with a study of the composition of tRNAs. Is there some systematic variation in the composition of tRNAs as a function of the kind of amino acid coded for by the tRNAs? For example, do the tRNAs for hydrophobic amino acids have more of the hydrophobic nucleotide, A, than average? Another, perhaps more direct question is, "Is the anticodon repeated in other parts of the tRNA?" The tRNA sequences available through 1985 have already been stored in our computer (floppy disk) and we have purchased a program for analyzing all tRNAs for composition, frequency of dinucleotide repeats, etc. In fact, most of this work has been completed but the accumulated data needs now to be analyzed. This work is rather straight forward.

Our major effort on tRNA evolution however, will center around an unusual and perhaps revolutionary idea, the possibility that there was early in evolution a universal tRNA i.e. a single RNA molecule that could serve as an adaptor for all amino acids, perhaps only recognizing them as classes of amino acids rather than as specific amino acids. Our candidate for that molecule is a precursor of today's 5s rRNA. There are several reasons for this selection. First, is the fact that E. coli tRNA^{Tyr} precursor and several other mature tRNAs

are highly homologous with E. coli 5s rRNA across the entire length of both molecules (Mullins et al., 1973). Secondly, the two ends of 5s rRNA base pair just as do tRNAs. Thirdly, 5s rRNA has in it regions which are quite homologous with the DHU and T Ψ C loops in tRNAs. These regions are fairly well conserved in tRNAs (Staves et al., 1987a). Fourthly, the 5s rRNA still is an integral part of protein synthesis though its function is not known. That 5s rRNA and tRNA might have a common origin is suggested by the fact that in eukaryotes both are synthesized by RNA polymerase III. Lastly, looking at the way the E. coli tRNA^{Tyr} is homologous with E. coli 5s rRNA suggests how the 5s rRNA (or a precursor) could give rise to many, if not all, of the tRNAs. In Fig. 9 we show the 5s rRNA could be processed to give mature E. coli tRNA^{Tyr}. Because the DHU and T Ψ C loops are necessary in all tRNAs, the beginning point on the circle could be the same for all tRNAs, but different anticodon loops and stems and variable loops could be obtained by selectively processing the intervening regions as shown.

Interestingly many of the known introns in tRNAs fall in the vicinity of the anticodon stem and loop and variable loop.

Preliminary work does show that the highest homology between tRNAs and 5s rRNAs is in the regions that represent the DHU and T Ψ C loops (Fig. 10). This suggests that now we look at the anticodon stem and loop and variable loop regions of tRNAs and look for homologies with the regions 100-120 and 1-17 of the 5s rRNA.

Evaluation of this will take some exploring and perhaps redirection after more data are obtained.

Laboratory equipment and space available

Beckman L-2 Ultracentrifuge

International Centrifuge

Laboratory Data Control HPLC apparatus

Waters HPLC apparatus

Nicolet 300 MHZ Spectrometer

Cary 219 UV-visible Spectrophotometer

JASCO ORD/CD UV5 Spectropotarimeter

Beckman DU Spectrometer

Beckman DBG Spectrometer

LKB fraction collector

ISCO fraction collector

Digital readout 3 decimal place pH meter, custom made

Various water baths (refrigerated)

Packard Radiochromatogram Scanner, Model 7201

Searle Scintillation Counter, Delta 300

BRL Vertical slab gel electrophoresis apparatus and power supply

Refrigerators and freezers

Labconco Freeze Dryer 5

Ample laboratory glassware, etc.

Cold room, 100 sq. ft.

Laboratory, 400 sq. ft.

Laboratory 160 sq. ft.

Laboratory 100 sq. ft.

Two offices, 100 sq. ft. each

References

- Crothers, D.M., Serro, T., and Soll, D.T., Proc. Natl. Acad. Sci, USA, 69 3063-3066 (1972).
- Goad, W.B. and Kanehisa, M.I., Nuc. Acid Res. 10 247-263 (1982).
- Gottikh, B.P., Krayevsky, A.A., Tarussova, N.B., Purygin, P.P. and Tsilevich, T.L., Tetrahedron, 26 4419-4433 (1970).
- Hecht, S.M., Tetrahedron, 33 1671-1696 (1977).
- Khaled, M.A., Mullins, D.W., Jr., Swindle, M. and Lacey, J.C., Jr., Orig. Life 13 87-96 (1983).
- Khaled, M.A., Mullins, D.W., Jr. and Lacey, J.C., Jr., J. Mol. Evol. 20 66-70 (1984).
- Lacey, J.C., Jr. and Mullins, D.W., Jr., Orig. Life 13 3-42 (1983).
- Lacey, J.C., Jr., Senaratne, N. and Mullins, D.W., Jr., Orig. Life 15 45-54 (1984a).
- Lacey, J.C., Jr., Hall, L.M. and Mullins, D.W., Jr., Orig. Life 16 69-79 (1985).
- Lacey, J.C., Jr., Mullins, D.W., Jr. and Watkins, C.L., J. Biomol. Struct. and Dynam. 3 783-793 (1986).
- Mullins, D.W., Jr., Lacey, J.C., Jr. and Hearn, R.A., Nature New Biol. 242 80-82 (1973).
- Mullins, D.W., Jr. and Lacey, J.C., Jr., J. Mol. Evol. 23 76-79 (1986).
- Profy, A.T. and Usher, D.A., J. Mol. Evol. 20 147-156 (1984a).
- Profy, A.T. and Usher, D.A., J. Am. Chem. Soc. 106 5030-5031 (1984b).
- Reuben, J. and Polk, F., J. Mol. Evol. 15 103-112 (1980).
- Schwartz, R.M., Gantt, M.S., and Dayhoff, M.O. in Atlas of Protein Sequence and Structure Vol. 5 suppl. 2 pp. 271-291. Nat. Biomed. Res. Found., Washington, D.C., 1976.
- Sprinzl, M. and Cramer, F., Prog. Nuc. Acid Res. and Mol. Biol. 22 1-69 (1979).
- Sprinzl, M. Moll, T., Meissner, F., and Hartmann, T., Nuc. Acid Res. 13 suppl. r1-r51 (1985).
- Sternglanz, H. Subramanian, E., Lacey, J.C., Jr., and Bugg, C.E., Biochem. 15 4797-4802 (1974).
- Staves, M.P., Bloch, D.P. and Lacey, J.C., Jr., Zeitschrift f. Naturforschung 42C 129-133 (1987a).

- Staves, M.P., Bloch, D.P. and Lacey, J.C., Jr., *Orig. Life*, in press (1987b).
- Taiji, M., Yokoyama, S. and Miyazawa, T., *Biochem.* 22 3220-3225 (1983).
- Taiji, M., Yokoyama, S. and Miyazawa, T., *Biochem.* 24 5776-5780 (1985).
- Usher, D.A. and Needels, M.C., *Adv. in Space Res.* 4 163-166 (1984c).
- Wagner, T., Cramer, F. and Sprinzl, M., *Biochem.* 21 1521-1529 (1982).
- Weber, A.L. and Lacey, J.C., Jr., *J. Mol. Evol.* 11 199-210 (1978).
- Weber, A.L. and Orgel, L.E., *J. Mol. Evol.* 16 1-10 (1980).
- Weber, A.L. and Orgel, L.E., *J. Mol. Evol.* 17 190-196 (1981).